SYSTEMS BIOLOGY OF *STAPHYLOCOCCUS AUREUS* INFECTION

EX VIVO AND *IN VITRO*

APPROVED BY SUPERVISORY COMMITTEE

First Name Last Name, credentials

James Thomas, MD

Damien Chaussabel, PhD

David Farrar, PhD

Felix Yarovinsky, PhD

SYSTEMS BIOLOGY OF *STAPHYLOCOCCUS AUREUS* INFECTION *EX VIVO* AND *IN VITRO*

by

ROMAIN BANCHEREAU

DISSERTATION / THESIS

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

May, 2012

ACKNOWLEDGEMENTS

The work presented herein could be broadly categorized as systems immunology, a multidisciplinary discipline involving cell biology, bioinformatics, biostatistics and medicine. As such, it involved many people that contributed in different ways to its completion. I would first like to thank my mentor, Dr. Octavio Ramilo, for giving me the opportunity to pursue research. Second, I would like to thank the four members of my Thesis Committee, Dr. James Thomas, Dr. David Farrar, Dr. Felix Yarovinsky and Dr. Damien Chaussabel for their continuous support and their critical evaluation of my progress over the years. I would like to thank the members of the UT Southwestern lab, including Dr. Monica Ardura and Dr. Asuncion Mejias, for the recruitment of patients, and the collection of clinical data, and Dr. Alejandro Jordan-Villegas, for the collection of clinical data and bacteriology work on clinical *S. aureus* isolates. Additionally, I would like to thank Evelyn Torres, RN and Dr. Juanita Lozano for their effort in patient enrollment. Of course, I would like to thank our patients and their families, for accepting to participate in our clinical studies.

All this research was conducted in close collaboration with the Baylor Institute for Immunology Research, where many people have spent time educating me at the bench or behind the computer screen. I first would like to thank its director, Dr. Jacques Banchereau, for providing scientific advice and motivational speeches, Dr. Karolina Palucka, for her guidance during the development of the dendritic cell transcriptional modules project, Dr. Virginia Pascual, for clinical and scientific input on the *S. aureus ex vivo* study, and Dr. Sangkon Oh for providing the original dendritic cell transcriptional

dataset that led to module development. I would like to thank all the members of the BIIR microarray core, led by Esperanza Anguiano, for training and support during microarray experiments, especially Phuong Nguyen, who processed all the samples for the whole blood *S. aureus* study, as well as Quynh Anh Nguyen and Benjamin Lemoine for chip scanning and data processing. A significant part of this work relied on advanced bioinformatics and biostatistics, for which I received tremendous support from Dr. Nicole Baldwin, Hui Xu, Dr. Jorge Rossello-Urgell and Dr. Derek Blankenship. In-house analytical tools were developed by Charlie Quinn, Yelena Hudson, Zhaohui Xu and Alexei Ionan. Additionally, I would like to thank former members of the BIIR flow cytometry core, who have trained me to use cytometers and analyze flow data including Dr. Tiziana Dipuccio, Adam Palazzo, Elizabeth Trahan and Sebastien Coquery. I would also like to thank the people involved in the *in vitro* dendritic cell work, including Dr. LuAnn Snipes and Dr. Shruti Athale for training me in the culture room and providing invaluable help during dendritic cell activation experiments, Yaming Xue for cultures of live flu virus, Dr. Yoichiro Ohne for phosphoflow collaboration and Dr. Jean-Pierre Gorvel for scientific advice.

Last but not least, I would like to thank all my friends and family members for their continuous support during these years of doctoral training.

SYSTEMS BIOLOGY OF *STAPHYLOCOCCUS AUREUS* INFECTION

EX VIVO AND *IN VITRO*

ROMAIN BANCHEREAU

The University of Texas Southwestern Medical Center at Dallas, 2012

OCTAVIO RAMILO, M.D.

Staphylococcus aureus has emerged as one of the most common community-acquired bacterial infections, with significant morbidity and mortality. Emergence of multidrug resistant strains worldwide, combined with limited treatment options demand novel approaches to further elucidate host-pathogen interactions, and especially host responses to infection. To this end, we leveraged systems biology approaches to better characterize the status of the host immune system during *S. aureus* infection *ex vivo* and *in vitro*.

The transcriptional profiles of PBMC and whole blood from patients with

community-acquired *S. aureus* infection were characterized by microarray analysis, and leukocyte population frequencies were measured by polychromatic flow cytometry. To refine our understanding of inflammatory networks involved, an *in vitro* system of antigen-presenting cell stimulation with various pathogens, including *S. aureus* as well as other bacteria and viruses, and their components, was used to identify early inflammatory programs induced in innate immune cells. To reduce the dimension and complexity of the data generated, we developed modular frameworks to analyze and interpret the fingerprints obtained from both the *ex vivo* and *in vitro* studies.

Overall, the blood transcriptional response to *S. aureus* infection was characterized by over-expression of innate immunity and hematopoiesis transcriptional programs, and under-expression of adaptive immunity programs. Flow cytometry and standard cell blood count (CBC) revealed an increase in absolute numbers of circulating monocytes, neutrophils and antigen-presenting cells, including dendritic cells and B cells, combined with a decrease in central memory T cells. To identify transcriptional correlates of clinical heterogeneity, we obtained individual fingerprints and derived the molecular distance to health, a numerical score of transcriptional perturbation for each patient. Patient-by-patient analysis without *a priori* knowledge of clinical diagnoses identified four major transcriptional clusters based on inflammation, erythropoiesis and interferon-induced profiles. Clinical presentation, bacterial dissemination and time between hospitalization and blood sampling were identified as major factors influencing the signature. The framework obtained from *in vitro* stimulation of monocyte-derived DC helped us refine the characterization of inflammatory programs activated during *S. aureus*

infection. In addition to inflammatory antibacterial programs, *S. aureus* induced a subset of interferon response modules, also observed in viral infections and autoimmunity, as well as a specific set of modules linked to cell compartmentalization and lipid biosynthesis.

Systems biology approaches provide a global and comprehensive assessment of host responses to acute bacterial infections, bringing a new understanding of disease pathogenesis and underlying patient heterogeneity.

TABLE OF CONTENTS

PRIOR PUBLICATIONS

1. **Banchereau R**, Jordan-Villegas A, Ardura MI, Mejias A, Baldwin N, Xu H, et al. Host Immune Profiles in *S. aureus* Infections Reflect the Heterogeneity of Clinical Manifestations. PLoS One. 2012.

2. Berry MP, Graham CM, McNab FW, Xu Z, Bloch SA, Oni T, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. Nature. 2010;466(7309):973-7.

3. Ardura MI, **Banchereau R**, Mejias A, Di Pucchio T, Glaser C, Allantaz F, et al. Enhanced monocyte response and decreased central memory T cells in children with invasive Staphylococcus aureus infections. PLoS One. 2009;4(5):e5446.

4. Klechevsky E, Liu M, Morita R, **Banchereau R**, Thompson-Snipes L, Palucka AK, et al. Understanding human myeloid dendritic cell subsets for the rational design of novel vaccines. Hum Immunol. 2009;70(5):281-8.

5. Keilbaugh SA, Shin ME, **Banchereau RF**, McVay LD, Boyko N, Artis D, et al. Activation of RegIIIbeta/gamma and interferon gamma expression in the intestinal tract of SCID mice: an innate response to bacterial colonisation of the gut. Gut. 2005;54(5):623-9.

LIST OF FIGURES

LIST OF TABLES

LIST OF PANELS

LIST OF APPENDICES

LIST OF ABBREVIATIONS

ASGPR – Asialoglycoprotein receptor

BDCA2 – Blood dendritic cell antigen 2

BIR – Baculoviral inhibitor of apoptosis repeat

CARD – Caspase activation and recruitment domain

CLR – C-type lectin-like receptor

DC – Dendritic cell

DC-SIGN – Dendritic cell-specific intercellular adhesion molecule-grabbing nonintegrin

DCIR – Dendritic cell immunoreceptor

GSEA – Gene set enrichment analysis

HIV – Human Immunodeficiency Virus

LC – Langerhans cell

LRR – Leucine-rich repeat

MINCLE – Macrophage-inducible C-type lectin

MMR – Macrophage mannose receptor

MRSA – Methicillin-resistant Staphylococcus aureus

MSSA – Methicillin-sensitive Staphylococcus aureus

NAIP5 – NLR apoptosis-inhibitory protein 5

NLR – Nod-like receptor

PAMP – Pathogen-associated molecular pattern

PBMC – Peripheral blood mononuclear cells

PRR – Pattern recognition receptor

PVCA – Principal variance component analysis

PYD – Pyrin domain

qPCR – quantitative polymerase chain reaction

RLR – RIG-I-like receptor

RT-PCR – real time polymerase chain reaction

SLE – Systemic lupus erythematosus

SoJIA – Systemic onset juvenile arthritis

TIR – Toll/Interleukin 1 receptor

TLR – Toll-like receptor

CHAPTER ONE: INTRODUCTION

Overview

The work presented herein relies on systems biology approaches to characterize the human immune response to the gram-positive bacterium *Staphylococcus aureus* (*S. aureus*). Using microarrays and polychromatic flow cytometry combined with data dimensionality-reducing analytical approaches, I first characterized the peripheral blood mononuclear cells (PBMC) of pediatric patients with acute *S. aureus* infection. In a second *ex vivo* study, I analyzed the whole blood immunological status of a new set of patients with *S. aureus* infection, focusing on how transcriptional patterns correlate with the heterogeneity of clinical presentations. Finally, to better understand the early transcriptional profiles elicited in innate immune cells by *S. aureus*, I extended my *ex vivo* observations to an *in vitro* system of antigen-presenting cells challenged with *S. aureus* and other pathogens.

Detection of pathogens by innate immune receptors

The innate immune system represents the first line of defense against invading pathogens such as viruses, bacteria, fungi and parasites. To induce specific protective host responses, innate immune cells including granulocytes, monocytes, macrophages,

dendritic cells, and B cells discriminate between these signals/pathogens through different families of phylogenetically-conserved innate receptors. These receptors detect pathogen-associated molecular patterns (PAMPs)[1] and are collectively referred to as pattern-recognition receptors (PRRs). They include membrane-bound receptors such as toll-like receptors (TLRs)[2-5] and C-type lectin-like receptors (CLRs)[6], and cytoplasmic receptors such as nucleotide-binding oligomerization domain (NOD)-like receptors (NLRs)[7], and more recently characterized helicases[8].

Membrane-bound receptors: Toll-like receptors (TLRs)

Discovered in the mid-1990s in *Drosophila*[9], TLRs are type I trans-membrane proteins with a Toll/Interleukin 1 receptor (TIR) endodomain and an exodomain containing leucine-rich repeats (LRR) which contains PAMP-binding sites. Ten functionally active TLRs have been identified so far in humans[10]. They can be broadly categorized into extracellular TLRs (TLR1, 2, 4, 5 and 6) and intracellular TLRs (TLR3, 7, 8 and 9). Extracellular TLRs recognize bacterial components, such as LPS (TLR4)[11,12], peptidoglycan (TLR2)[13], lipoteichoic acid (TLR2)[14] and flagellin (TLR5)[15]. Intracellular TLRs on the other hand detect single and double-stranded nucleic acids from viruses and bacteria. TLR3 recognizes double-stranded RNA from reoviruses[16], while TLR7 and TLR8 recognize single-stranded RNA from viruses such as respiratory syncytial virus, influenza A or human immunodeficiency virus (HIV)[17,18]. Finally, TLR9 recognizes CpG DNA motifs present in both viruses and bacteria[19]. The ligand for human TLR10 has not been identified yet. PAMP recognition by TLRs is summarized in Figure 1.

TLR expression varies significantly between leukocyte populations, thereby conferring them specific capacities to respond to various pathogens. For example, plasmacytoid dendritic cells (pDC) express high levels of TLR7 and produce large amounts of type I interferons upon viral single-stranded RNA sensing[20]. Conversely, myeloid dendritic cells (mDC) and macrophages express high levels of TLR2 and TLR4 and induce pro-inflammatory responses in response to bacterial stimuli[21]. TLR8 is highly expressed in monocytes[22]. TLR2 has also been reported on T cells[23], which allows them to respond to lipopeptides in an antigen-presenting-cell independent fashion.

Membrane-bound receptors: C-type lectin-like receptors (CLRs)

C-type lectin-like receptors (CLRs) represent the second major family of known membrane-bound PRRs. These include the mannose receptor MMR[24], DEC205[25], DC-SIGN[26], BDCA2[27], Dectin-1[28], DCIR[29], Lox-1[30], Langerin[31], MINCLE[32] and ASGPR[33]. In addition to their role in leukocyte trafficking and APC-T cell interactions, CLRs can detect and take up antigen in lysosomal compartments for processing and presentation by APCs[6]. CLRs recognize mannose, fucose and glucan carbohydrates through their Ca^{2+} -dependent carbohydrate-recognition domains (CRDs)[6,34], which allow them to detect a variety of pathogens. DC-SIGN detects mycobacteria, HIV-1[35], *Candida albicans*, and *Leishmania*. Langerin detects *Mycobacterium leprae* and HIV-1. The beta-glucan receptor Dectin-1 binds and internalizes fungi[36].

In contrast to TLRs, CLRs do not seem to discriminate between self and non-self, and can also recognize carbohydrate residues on self-glycoproteins to induce self-

tolerance[6]. In addition, recent observations suggest that CLRs can directly activate downstream signaling cascades upon glycan detection through Syk and Raf1-dependent pathways[37]. Similarly to TLRs, CLR expression on the surface of antigen-presenting cells varies. Langerin is mainly expressed on Langerhans cells (LC)[31] and a dermal DC subset. MINCLE seems to be specific to macrophages[32]. BDCA-2 is restricted to pDC[27], while DCIR seems to be preferentially expressed in CD14+ CD34-derived, dermal dendritic cells[6] and T cells upon HIV-1 infection.[38] While MMR and DC-SIGN are abundantly expressed on the surface of immature monocyte-derived DC, they are not readily detected on the surface of fresh blood DC or LC[6]. Intense efforts are under way to harness the specificity of expression of CLRs on APC subsets to develop reagents for *in vivo* targeting of vaccines components[37-39].

Several studies suggest an important collaboration between TLRs and CLRs in the regulation of the innate response to pathogens. BDCA-2, which recognizes complex sugars such as asialo galactosyl-oligosaccharides[40], modulates pDC responses to TLR9-activating viral nucleic acids by inhibiting the production of type I interferons[41]. Dectin-1 in combination with TLR2 or TLR4 is involved in the recognition of *C. albicans*[42]. DCIR internalization was shown to inhibit TLR8-mediated cytokine production in human monocyte-derived DC[43]. The synergy between PAMP receptors is only partially understood, and the study of positive and negative feedback interactions between the downstream signaling pathways involved will require approaches that can measure all these parameters quantitatively and qualitatively over time.

Cytoplasmic Bacterial Receptors: NOD-like receptors (NLRs)

TLRs and CLRs are membrane receptors that detect pathogens on the cell surface or engulfed in vacuoles such as the early and late endosomes. However, many pathogens can avoid or subvert these membrane detection mechanisms and spread to the cytosol. To detect and eliminate these threats once they crossed cell membranes, the innate immune system relies on two major families of cytoplasmic PRRs: NLRs and helicases.

Mammalian NLRs comprise more than 20 members, which are characterized by a C-terminal leucine-rich repeat (LRR) domain, a central nucleotide binding NACHT domain and an N-terminal protein-protein interaction domain composed of either a caspase activation and recruitment domain (CARD) or a pyrin domain (PYD)[44]. They can be further divided into non-inflammasome-activating NLRs (Nod1, Nod2 and NLRX1) and inflammasome activating (NLRP1, NLRP3 and NAIP).

Nod proteins belong to the NLR family and detect distinct bacterial peptidoglycan (PGN) structures. Nod1 detects *meso*-diaminopimelic acid (DAP)-containing PGN[45,46] which is more commonly found in gram-negative bacteria. Nod2 detects muramyl dipeptide (MDP)[47], which is common to gram-negative and gram-positive bacteria. Upon antigen ligation, Nod1 and Nod2 propagate the signal through their CARD domain to RIP2[48] and activate NF-kappa B[49,50]. Nod1 is essential in the detection of and protection against the bacterium *Helicobacter pylori*[51,52] and the parasite *Trypanosoma cruzi*[53]. Nod2 is involved in the detection of *Mycobacterium tuberculosis*[54], *Listeria monocytogenes*[55], *Toxoplasma gondii*[56] and viral single-stranded RNA[57]. In addition to the Nods, NLRX1, which contains a CARD-related 'X' domain, is the only known NLR family member targeted to the mitochondria. There, it interacts with

mitochondrial anti-viral signaling protein (MAVS) to mediate antiviral responses[58]. NLRX1 was also shown to induce reactive oxygen species (ROS)[59], which amplify proinflammatory responses against bacteria[60].

The second major NLR subfamily includes the NLRPs. NLRP1 and NLRP3 signal through their PYD domain and activate the inflammasome[61], which results in the processing and activation of proinflammatory cytokines IL1-beta and IL18[62] by caspase 1 (CASP1). Studies have shown that NLRP3 can be directly triggered by bacterial components such as MDP[63], but also indirectly by danger signal such as ATP, toxins or intracellular potassium depletion[64]. Whole pathogen studies have shown that the NLRP3 inflammasome can be triggered by the fungi *Candida albicans* and *Saccharomyces cerevisiae*[65], the gram-positive bacteria *Listeria monocytogenes* and *Staphylococcus aureus*[64] and the viruses Sendai virus, adenovirus and influenza[66,67]. The NLRP1 inflammasome is involved in detection of *Bacillus anthracis*[68] and MDP[69], but its exact mechanism of activation remains elusive.

The third major NLR subfamily includes IPAF and NAIP5. IPAF contains a CARD domain, through which it signals directly to CASP1[70]. The IPAF inflammasome is primarily involved in the detection of gram-negative bacteria including *Salmonella typhimurium*[71], *Shigella flexneri*[72], *Legionella pneumophila*[73], and *Pseudomonas aeruginosa*[74,75]. NAIP5, a structurally unique NLR with an N-terminal set of baculoviral inhibitor of apoptosis repeat (BIR) motifs, senses flagellin in a TLR5-independent fashion. NAIP5 is involved in the detection of *Legionella pneumophila*[76-79].

Cytoplasmic Viral Receptors: Helicases

NLRs are mainly intracellular bacterial detectors. Viruses also cross cell membranes and invade the cytosol, where they release nucleid acids. Viral PAMPs such as ssRNA, dsRNA and dsDNA are detected by a family of helicases known as RIG-I-like receptors (RLRs), including retinoic acid-inducible gene (RIG-I)[80] and melanoma differentiation associated gene 5 (MDA5)[81]. These helicases induce signaling cascades involving interferon regulatory factors IRF3 and IRF7, leading to the production of type I interferon IFN-beta, which is a major component of the antiviral response[82].

RIG-I contains a DExD/H RNA helicase domain and two effector CARD domains on its amino-terminus. RIG-I was originally found to bind dsRNA (poly(I:C)) but not ssRNA (poly(A)) or dsDNA[80]. Subsequent studies showed that RIG-I can recognize single-stranded RNA molecules with a 5'-phosphates end[83,84]. Thus, RIG-I detects RNA from viruses of the paramyxovirus (e.g. measles virus) and orthomyxovirus (e.g. influenza virus) families.

While MDA5 is structurally similar to RIG-I[82], studies have shown that it is critical for the recognition of picornaviruses[85,86] (e.g. enteroviruses, rhinoviruses), but not paramyxoviruses or influenza viruses. This specificify was explained by the absence of phosphorylation on the 5' end of ssRNA from these viruses. In addition, two new helicases, DHX36 and DHX9, have been identified as dsDNA cytosolic sensors in pDC[8], which consequently produce type I interferons through IRF7 in a TLR9-independent fashion.

Complex connections between innate immune networks

In summary, the innate immune system has developed an array of membrane (Figure 1) and cytoplasmic (Figure 2) receptors with various ligand specificities, which, alone or in synergy, induce the activation of complex signaling cascades that result in the production of inflammatory mediators and the activation of adaptive immunity.

Two major downstream transcriptional axes have been characterized in depth: i) the TRIF and IRF-mediated IFN signaling pathway, which leads to the production of type I & type III interferons, and the MyD88-mediated NF-kappa B signaling pathway, which leads to the production of inflammatory cytokines IL1-beta, IL6, IL8 and TNF-alpha. The IFN signaling pathway is known to play a major role in response to viral infection, while the NF-kappa B pathway seems to be essential for anti-bacterial activity and clearance of extra-cellular pathogens.

There is growing evidence that these two signaling cascades are intertwined and may negatively regulate each other. This is supported by clinical observations that patients with secondary bacterial infection with *S. aureus,* usually following a flu episode, have a much worse prognosis than patients with *S. aureus* infection alone. *In vivo* studies have suggested a role of IFN-induced glucocorticoids in the dampening of the antibacterial response[87]. Furthermore, *in vitro* studies have revealed inhibition of inflammasome activation by type I IFNs in a STAT1-dependent mechanism, suggesting cell-intrinsic negative feedback[88]. As the complexity of these innate signaling networks gets unraveled, it is becoming increasingly difficult to predict the effects of enhancing or inhibiting these pathways on the biological system under study as a whole. In the last decade or so, new approaches have been developed that aim at characterizing systemic responses to specific conditions at the transcriptional, translational and epigenetic levels

in an unbiased fashion. Together, these approaches formed an emerging discipline known as systems biology.

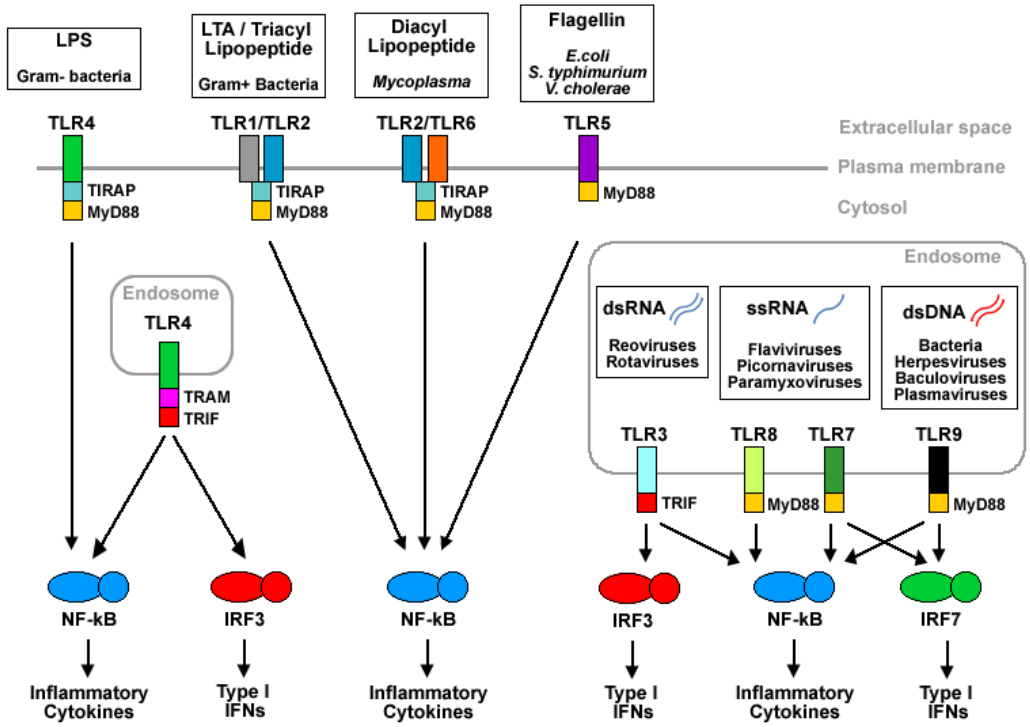


Figure 1: Detection of PAMPs by toll-like receptors

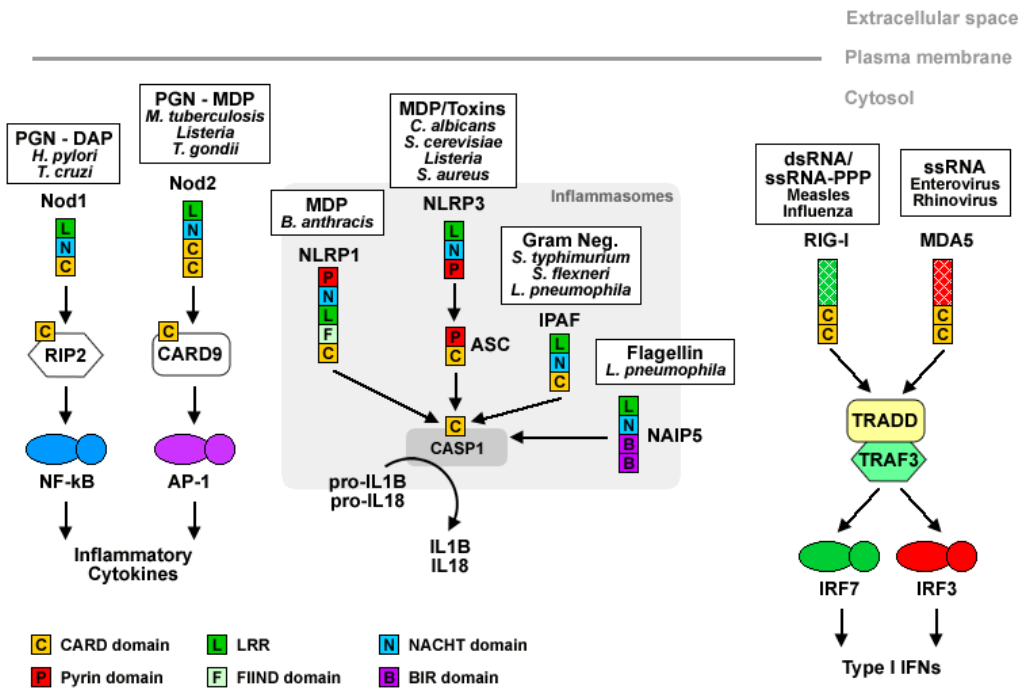


Figure 2: Detection of PAMPs by cytosolic PRRs

Systems biology sheds light on the regulation of complex molecular networks in health and disease

Considering the system as a whole

In recent years, evidence has accumulated to suggest that the signaling cascades triggered by the various families of PRRs involved in response to infection interact to form complex regulatory networks, with multiple positive and negative feedback mechanisms. While essential to dissect specific signaling pathways and identify key molecules involved, traditional single knock-out approaches in animal models have proven limited to understand how these individual components integrate into the global immune response to infection *in vivo* in human. This complexity gave birth to the discipline of systems biology, which the Institute for Systems Biology (Seattle, WA) defines as "the study of an organism, viewed as an integrated and interacting network of genes, proteins and biochemical reactions which give rise to life". Unlike traditional hypothesis-driven scientists, systems biologists use high-throughput technologies, such as microarray, deep-sequencing, Chromatin Immuno-Precipitation Sequencing (ChIP-SEQ), polychromatic flow cytometry and computational modeling to probe elements of a complex system in an unbiased fashion, generating hypotheses along the way.

Emergence of high-throughput technologies to quantify nucleic acids

In the late 1980s an early 1990s, investigators started developing RT-PCR-based approaches to quantify nucleid acids on a larger scale, with the goal of interrogating thousands of transcripts at a time. Among the first arrays available, spotted arrays were

obtained by dropping pre-synthesized cDNA oligonucleotides on filter papers or on miniaturized coated glass slides with a pin-spotting device[89]. Private companies such as Affymetrix started producing oligonucleotides microarrays on an industrial scale, using photolithographic synthesis, a process whereby 25-mer oligonucleotide sequences are built on the array using light-sensitive masking agents[90]. In contrast, Illumina developed "beadchip" arrays consisting of shorter cDNA probes attached to beads coated randomly in higher density on glass plates, which are cheaper to manufacture.

Variations in mRNA transcript abundance between samples are detected differently based on the microarray type. Two-color microarrays rely on the mRNA from two samples to be color-coded with two different dyes (usually red and green) and recombined into one mixture, which is then hybridized on chips. Based on the abundance of a particular transcript in each sample, the detected color for a particular spot or transcript will vary. This pairwise comparison approach was well suited for small numbers of samples, but not scalable. With one-color microarrays, such as Affymetrix GeneChips or Illumina BeadArrays, which were used throughout this work, each sample requires its own hybridization, which allows the investigator to compare hundreds of samples at a time. However, only relative abundance of a particular transcript as compared to other transcripts from the same sample can be quantified. For this reason, microarray-based observations are often validated by truly quantitative assays such as quantitative PCR (qPCR)[91].

In the past 10 years, new technologies have emerged that combine the high-throughput of microarrays with the absolute quantification readout and sensitivity of qPCR. These include the Nanostring[92], that uses color-coded molecular barcodes that

hybridize to target mRNA transcripts, and the OpenArray, which is a high-throughput Taqman RT-PCR platform developed by Applied Biosystems. Additionally, sequencers that can not only quantify absolute mRNA expression but also interrogate a specific transcript's splicing variants or identify single nucleotide polymorphisms (SNPs) in a single run in an increasingly cost-effective way are quickly replacing traditional microarray platforms.

Application of microarrays to the analysis of innate response to pathogen in vitro

With the cost of gene expression microarray experiments gradually dropping and their sensitivity and reproducibility increasing, their use increased for a broad range of clinical and basic research applications. While 55 publications reported the use of microarrays between 1995 and 1998, 5400 studies using microarrays were published in 2005 alone[93]. A major application of microarrays is the identification of transcriptional programs induced in immune cells in response to pathogen and other stress stimuli, which was reviewed in details by Jenner and Young in 2005[94].

Early studies identified interferon-induced genes in a human fibrosarcoma cell line[95] and transcriptional responses to human cytomegalovirus (HCMV) in human fibroblasts[96]. Subsequent studies focused on the transcriptional profiles of human immune cells, including peripheral blood mononuclear cells (PBMC)[97], whole blood[98], neutrophils[99,100], monocytes[101,102], blood-sorted myeloid[103] and plasmacytoid dendritic cells[104], monocyte-derived dendritic cells (IL4/GM-CSF DC)[105,106], macrophages[106-108], epithelial[109,110] and endothelial[111,112] cells in response to a variety of bacterial, fungal and

viral pathogens including *E. coli*, *S. enterica*, *S. aureus*, *S. pneumoniae*, *M. tuberculosis*, *N. gonorrhoeae*, *C. trachomatis*, *B. pertussis*, *C. albicans*, Influenza A, Herpes implex virus (HSV), Hepatitis B and C, or human immunodeficiency virus (HIV). They identified common transcriptional responses, including IFN-induced genes (IFI family, OAS family, CXCL10), inflammation mediators (IL1, IL6, IL8, TNF-alpha), immune response activators and regulators in response to these pathogens.

Comparison of responses to a single stimulus between different cell populations identified population-specific responses, largely dependent on their expression or PRRs, highlighting the diversity between innate immunity cells[106]. These cell-specific transcriptional programs are at the center of deconvolution studies, which aim at characterizing the cellular composition of heterogeneous tissues based on their combined mRNA fingerprint[113]. Conversely, studies comparing transcriptional responses of a single cell type to multiple stimuli helped characterize TLR-dependent and independent pathways involved in response to PAMPs. Transcriptional profiling further helped to understand how viruses and bacteria modulate host cell transcription to survive, expand and induce disease.

More recently, systems biology approaches such as Nanostring combined with RNA interference have been used to identify transcriptional regulators of the IFN/antiviral and pro-inflammatory/antibacterial pathways in mouse bone-marrow-derived dendritic cells[114]. These studies support the existence of complex feedback mechanisms between these pathways, and that type I IFN may inhibit pro-inflammatory responses, in particular inflammasome-mediated IL-1 release by innate immune cells[88].

Ex vivo application of microarrays to understand human disease

While transcriptional profiles induced in various human cell populations after *in vitro* challenge with a variety of pathogens provide information about signaling cascades, cell-specific responses and immune system evasion mechanisms used by these pathogen, they are not representative of *in vivo* conditions. The use of microarrays to characterize patient transcriptional profiles *ex vivo*, when cells are lysed shortly after tissue sampling, has proven useful in the diagnosis, prognosis and understanding of disease pathogenesis in autoimmunity, cancer or infectious diseases. Transcriptional profiling studies in human are however limited by the difficulty to collect tissues such as the spleen, lymph nodes, liver or brain. Circulating leukocytes, while trafficking to and from the sites of infection and inflammation, represent an accessible source of molecular information, which can be assessed by genome-wide analysis of whole blood[115].

Blood transcriptional profiling has been used extensively to characterize autoimmune diseases. It identified the characteristic type I-IFN signature in systemic lupus erythematosus (SLE)[116,117], and pinpointed the IL1-beta signature in systemic onset juvenile idiopathic arthritis (SoJIA)[118], leading to diagnostic[119,120] and therapeutic advances with the IL-1R blocker Anakinra. Other autoimmune and autoinflammatory diseases studied through blood or PBMC transcriptional fingerprinting include psoriasis[121], type I[122] and II[123] diabetes, multiple sclerosis[124], rheumatoid arthritis[125], Sjogren's syndrome[126] and vasculitis[127]. These studies were recently reviewed by Pascual et al[128] and Chaussabel et al[115].

In the context of infectious diseases, blood transcriptional profiling has been used to characterize the interaction of pathogens with the immune system as well as the quantity and quality of the immune response elicited against them. Blood signatures have been described in the context of acute local infections, such as upper respiratory tract infections[129], urinary tract infections and skin abscesses[130], acute systemic infections, such as sepsis[131,132] and influenza[129], or chronic infections, such as tuberculosis[133], HIV[134-137] or HCV[138,139]. Microarrays have been further used to discriminate between patients acutely infected with distinct classes of pathogens[130], to separate responders and non-responders to treatment in HCV infection[139], and to predict disease outcome during acute febrile stage of dengue infection[140]. These studies in autoimmune and infectious diseases suggest that microarrays could potentially be used in the clinic for diagnosis, prognosis and prediction of response to specific treatments.

Simplifying microarray readouts: analytical approaches to reduce data dimensionality

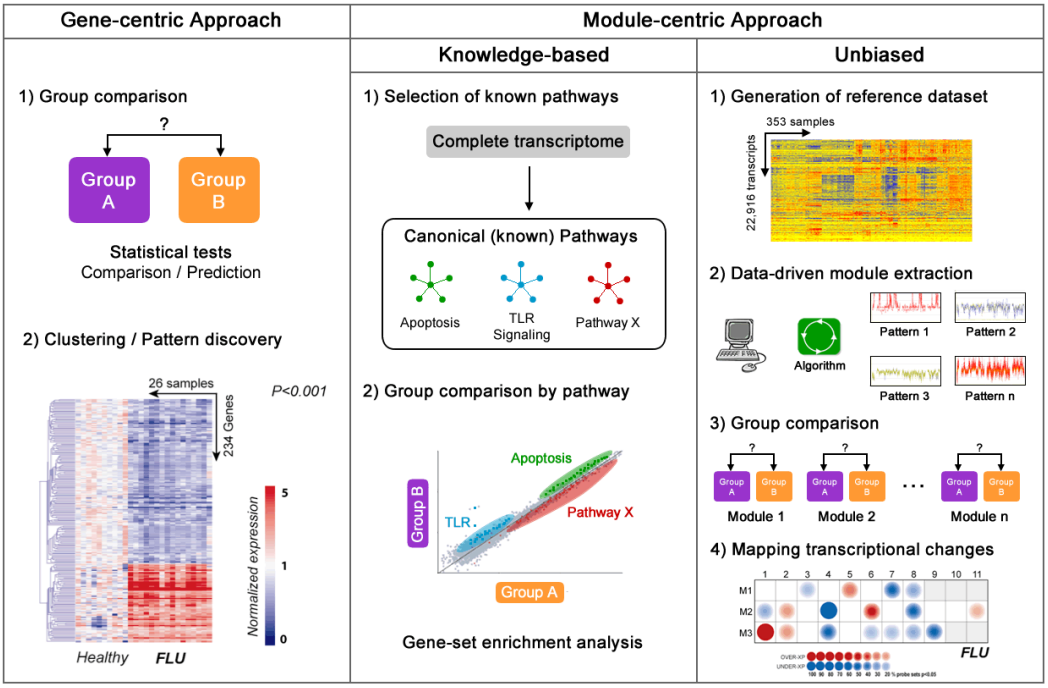
A major issue in microarray analysis is the high dimensionality of the data generated, even from small studies. A study including 10 patients, 10 healthy controls and 5 time points per subject will generate approximately 5 million data points if 50,000 transcripts are analyzed per sample. In addition, gene-centric microarray analyses, including clustering, and manual annotation of lists obtained through statistical filters (T-test, ANOVA) tend to i) ignore transcripts that do not pass arbitrarily-set thresholds, or are less studied in the context of the biological question addressed; ii) suffer from elevated signal-to-noise ratio; iii) simply list genes associated with a phenotype without thorough understanding of the molecular mechanisms involved[141]. Thus, there is a need

for analytical approaches that reduce data dimensionality and facilitate the visualization and functional interpretation of these large datasets.

To address the issues stemming from individual gene analysis, several groups have developed methods that consider groups of genes, or modules, as the basic building blocks for transcriptional profiling. Different methods, summarized in Figure 3, have been used to derive these modules. Knowledge-based classification of genes into pathways is used in gene-set enrichment analysis (GSEA)[142], to identify known pathways differentially regulated between conditions. This approach allows the detection of small changes occurring in functionally related transcripts, which are not detected using gene-centric statitistical filters[142,143]. Another method to derive gene modules consists in identifying gene sets with similar behavior across a dataset of reference, without the bias of knowledge-based classification. Segal et al. applied this concept to develop the 'cancer module map' by identifying sets of genes coordinately regulated across a reference dataset of 1,975 microarrays spanning 22 tumor types[144]. These cancer modules allowed them to identify transcriptional commonalities and differences between different tumors.

Following a similar unbiased data-driven approach, Chaussabel et al. developed the first set of PBMC transcriptional modules in immune-related diseases, compiling datasets spanning autoimmune diseases, infectious diseases and transplantation[145]. These modules were used in the context of SLE to define the global signature of the disease, to derive a multivariate transcriptional score of disease progression and to monitor the onset of SLE flares. They were subsequently applied to SoJIA[146] and infectious diseases such as tuberculosis[133] and septicemic melioidosis[131]. The work presented herein utilizes these modules, as well as 2 new sets of modules specifically designed for the analysis of whole

blood and dendritic cells transcriptomes, to systematically characterize the human immune transcriptional profiles induced *ex vivo* and *in vitro* by the gram-positive bacterium *Staphylococcus aureus*.



Adapted from Chaussabel et al., Immunity, 2008 and Segal et al., Nature Genetics, 2005

Figure 3: Three approaches used for microarray analysis

Left panel: gene-centric approach. This traditional method to identify transcripts differentially regulated between groups of samples involves statistical tests such as t-test for 2 samples or ANOVA for more than 2 samples as well as class prediction algorithms that identify genes lists that best discriminate between the tested groups. **Center panel: knowledge-based module-centric approach.** This approach consists in analyzing genes organized in groups (or modules) based on known pathways / interactions. This permits the identification of small changes between large groups of genes that would be missed with traditional gene-centric comparisons. **Right panel: unbiased module-centric approach.** This method makes no assumption about known connectivity between transcripts. It first identifies clusters (modules) of transcripts co-regulated across a large

reference dataset using clustering algorithms such as the k-means algorithm. Comparisons between groups of interests are then carried on a module-by-module basis. The results of these comparisons are displayed with heatmaps or modular fingerprints that represent over or under-expression of each module for the condition/group represented.

***Staphylococcus aureus*: a re-emerging threat to humans**

A threat to the population

First described in 1880 by Alexander Ogston[147], *Staphylococcus aureus* (*S. aureus*) remains a leading cause of invasive bacterial infections worldwide, resulting in significant morbidity and mortality. An asymptomatic skin and nostril colonizer in up to 30% of healthy individuals, *S. aureus* can become a dangerous pathogen when it breaches the physical skin and mucosal barriers. In the United States alone, an estimated 94,360 cases of methicillin-resistant *S. aureus* (MRSA) were reported in 2005, leading to 18,650 deaths[148], which is greater than those caused by HIV infection. Additionally, the economic burden of *S. aureus* infections for hospitals in terms of increased length of stay and medication was estimated to be $14.5 billion in 2003[149].

Changing Epidemiology

Originally susceptible to beta-lactam antibiotics such as penicillin[150], *S. aureus* gradually developed resistance to most antibiotics available, probably through horizontal gene transfer[151]. Shortly after the introduction of penicillin in the clinic in 1940, penicillin resistant strains of the lineage phage 80/81 were isolated[152]. They produced beta-lactamase, an enzyme that prevents the antimicrobial effect of these antibiotics by disrupting their beta-lactam ring. The introduction of methicillin, a beta-lactam antibiotic insensitive to beta-lactamases, on the market in 1960, aimed at targeting methicillin-sensitive *S. aureus* (MSSA) strains. It was however followed within a year by the

isolation of the first methicillin-resistant *S. aureus* strains (MRSA)[153]. The increasing burden of MRSA infection in the next 30 years in hospital settings led to the use of vancomycin, the only available antibiotic reliably effective against MRSA. Not surprisingly, this led to the emergence of vancomycin-intermediate *S. aureus* (VISA)[154] in 1997 and vancomycin resistant *S. aureus* (VRSA) in 2003[155]. Similarly, the alternative use of the synthetic antibiotic linezolid led to the identification of the first outbreak of linezolid-resistant *S. aureus* (LRSA) in Spain in 2008[156].

Antibiotic resistance is not the only evolving feature in *S. aureus* epidemiology. Historically a nosocomial pathogen, mainly affecting surgical and immuno-compromised patients in hospital settings, *S. aureus* has recently emerged as a community-acquired (CA) threat and has become a major cause of bacterial infection in otherwise healthy children and adults[157-161]. While CA *S. aureus* infections were mostly caused by MSSA[162], this distribution has changed in the United States in the past ten years, with the rapid spread of CA-MRSA strains USA400 followed by USA300[162,163]. The rapidity and extent of the spread of CA-MRSA worldwide, combined with high virulence leading to tissue-destructive infections such as necrotizing fasciitis[164] and necrotizing pneumonia[165] has raised major concerns about the acquisition of antibiotic resistance by these community strains.

Clinical presentations

Patients with CA *S. aureus* infection present with a wide spectrum of clinical illness, ranging from mild soft tissue infections to invasive disease such as bacteremia[166],

pneumonia, musculosqueletal infections[167] (including osteomyelitis, suppurative arthritis and pyomyositis), endocarditis, meningitis, and disseminated disease, including sepsis and toxic shock syndrome. This ability of *S. aureus* to disseminate through the blood stream to multiple organs, including bones, joints, kidneys and lungs highlights the versatility of this bacterium and commands the development of an array of therapeutic treatments that target various infection foci.

Interaction with the immune system

*S. aureu*s expresses an array of virulence factors that affect the host immune system in various ways. The single-layered wall of this gram-positive coccal bacterium is mainly composed of lipoteichoic acid (LTA) and peptidoglycan (PGN), which are detected by TLR2, CD14 or peptidoglycan recognition proteins (PGLYRPs)[168] on the cell surface and NOD2 in the cytoplasm. Both structural components were shown to induce TNF-alpha, IL6 and IL10 production in human monocytes and T cells[169]. Monocytes additionally secrete the pro-inflammatory cytokines IL1 and IL8 in response to *S. aureus* wall components.

S. aureus produces multiple toxins including nine staphylococcal enterotoxins SEA, SEB, SEC, SED, SEE, SEG, SEH, SEI and SEJ[170], toxic-shock syndrome toxin TSST-1[171], which all behave as super-antigens by locking the MHC Class II / T cell receptor interface. Through this mechanism, they induce massive release of pro-inflammatory cytokines by macrophages and T cells, leading to high fever, shock, capillary leak and multiorgan dysfunction[172]. Additionally, most bacterial clones produce

alpha-haemolysin, which lyses epithelial cells, erythrocytes, fibroblasts and monocytes, leading to hypotension, thrombocytopenia and reduced oxygenation, all symptoms of sepsis[173].

Recent clones such as the prevalent MRSA USA300, as well as the penicillin-resistant phage type 80/81 clone responsible for serious outbreaks in the 1950s produce Panton-Valentine leukocidin (PVL)[174], which displays leukocyte-killing activity and may play a major role in necrotizing pneumonia[175]. Finally, the recently discovered alpha-type phenol-soluble modulins (PSMalphas), which are short peptides highly expressed in CA-MRSA[176], recruit and lyse neutrophils, thereby promoting the necrotic process and *S. aureus* pathogenesis.

Current treatments

Antibiotics remain the drug of choice against *S. aureus*. Penicillin is still used for beta-lactamase non-producing strains, although these are rarely observed nowadays. Semi-synthetic penicillins such as oxacillin or nafcillin are used for beta-lactamase-producing strains. While vancomycin is the preferred agent against MRSA strains, its effects are limited by recurrent bacteremia during treatment[177], potential nephrotoxicity[178], and its relative ineffectiveness against bacterial pneumonia due to low lung penetration[179]. In such cases, more recent drugs such as linezolid or daptomycin can be alternatively used, with comparable efficacy. In addition to antimicrobial therapies, draining of suppurative sites and removal of accessible infection foci are common procedures to prevent further dissemination of infection.

To this day, there is no commercially available preventive vaccine against *S. aureus* and the presence of antistaphylococcal antibody titers in previously infected individuals do not seem to correlate with protection against subsequent reinfection[180]. Several anti-staphylococcal polyclonal and monoclonal antibodies are currently under development, but their efficacy remains to be determined. The rapidly changing epidemiology of *S. aureus*, including acquisition of drug resistance and novel virulence factors, combined with the complexity of host immune responses elicited through an array of PAMP receptors during infection, provide an adequate context for the use of systems biology to study *S. aureus* pathogenicity.

Goal, hypothesis and aims

Goal

The goal of the work presented herein is to apply systems biology approaches, including microarrays, polychromatic flow cytometry and data dimensionality-reducing analytical approaches to better characterize the human immune response to *S. aureus* infection, *ex vivo* and *in vitro*.

Hypothesis

I hypothesized that using systems biology approaches such as microarrays and polychromatic flow cytometry to study the transcriptional and phenotypical immunological status of circulating human leukocytes *ex vivo* and *in vitro* would help us better understand host-pathogen interactions during *S. aureus* infection.

Aims

Aim #1: To characterize the global transcriptional signature of peripheral blood mononuclear cells (PBMC) from pediatric patients with community-acquired *S. aureus* infection using gene and module-centric approaches, and correlate molecular profiles obtained with leukocyte subpopulations distribution as determined by polychromatic flow cytometry.

Aim #2: To expand the PBMC study to whole blood samples, and explore transcriptional correlates of clinical heterogeneity observed amongst patients using a new framework of modules designed for analysis of whole blood fingerprints.

Aim #3: To develop an alternate analytical modular framework to study early transcriptional responses of antigen-presenting cells to infectious pathogens *in vitro*, and apply it to characterize the molecular fingerprint induced by *S. aureus in vitro*.

CHAPTER 2: METHODOLOGY

Ethics Statement

The studies described herein were conducted according to the principles expressed in the Declaration of Helsinki. The studies were approved by the Institutional Review Boards of the University of Texas Southwestern Medical Center and Children's Medical Center of Dallas (IRB #0802-447) and Baylor Institute of Immunology Research (BIIR, IRB # 002-141). Informed consent was obtained from legal guardians and informed assent was obtained from patients 10 years of age and older prior to any study-related procedure.

Sample collection

PBMC isolation

Blood samples (3–8 mL) were collected in acid-citrate-dextrose tubes (ACD tubes, BD Vacutainer, Franklin Lakes, NJ) and delivered to the laboratory at room temperature for processing. Peripheral blood mononuclear cells (PBMCs) were isolated by density gradient centrifugation using Ficoll-hypaque technique and lysed in RLT reagent (Qiagen, Valencia, CA) with b-mercaptoethanol (BME) and stored at -80uC until processing. Samples were run in batches by the same laboratory team to ensure standardization of quality and handling of samples. Total RNA was isolated using the RNeasy Mini Kit (Qiagen, Valencia, CA) per the manufacturer's instructions and RNA integrity was assessed using an Agilent 2100 Bioanalyzer (Agilent, Palo Alto, CA).

Patient Information and Classification

PBMC study – Patient information

Blood samples were collected from 77 pediatric patients, including 53 patients with *S. aureus* infection and age, gender and ethnicity-matched 24 healthy controls. Children with suspected or proven polymicrobial infections, underlying chronic disease, immunodeficiency, or those who received steroids or other immunomodulatory therapies were excluded. Control samples were obtained from healthy children undergoing elective surgical procedures and at healthy outpatient clinic visits. Nasopharyngeal viral cultures were obtained in both patients and controls to exclude viral co-infections. Children hospitalized with acute *S. aureus* infections were offered participation in the study after microbiologic confirmation of the diagnosis by standard bacterial culture of blood or tissue specimens. Patients were analyzed in 3 groups: training (20 *S. aureus*, 10 healthy controls), test (22 *S. aureus*, 10 healthy controls), and validation sets (11 *S. aureus*, 14 healthy controls). The design of this study is summarized in Panel 1.

Whole blood study – Patient Information

Blood samples from 99 patients hospitalized with community-acquired *S. aureus* infection and 44 healthy controls were collected in tempus tubes (Applied Biosystems, PN 4342792). Patients represented the clinical spectrum of acute *S. aureus* infections, including skin and soft tissue abscesses, bacteremia, osteomyelitis, suppurative arthritis, pyomyositis, pneumonia with empyema, and disseminated disease. Patients with a

diagnosis of toxic shock syndrome, polymicrobial infections, or treated with corticosteroids in the preceding four weeks were excluded. Viral direct fluorescent antibody testing and/or culture of the nasopharynx were performed in all patients and healthy controls to exclude concomitant viral infections. Patient demographic data and clinical characteristics are summarized in Table 7 and Table 8. The median duration of hospitalization was ten days (range: 1-98 days). The median time from patient hospitalization to blood sample acquisition was five days (range: 1-35 days).

Whole blood study – Patient classification

The study cohort of 99 patients and 44 healthy controls was divided into independent training (40 patients, 22 healthy controls) and test sets (59 patients, 22 healthy controls) (Table 10). Patients were categorized according to two schemes (Table 11) based on assessment by an independent clinician who was blinded to the transcriptional data: i) by localization of infection, defined as local (n=10), invasive (n=74) or disseminated (n=13); ii) by clinical presentation, separating patients with skin and soft tissue abscesses with negative blood culture (n=10), patients with osteoarticular infections (n=56), and patients with pneumonia (n=11).

Whole blood study – Draw index and hospitalization quarter

This cross-sectional study included samples drawn at different days during hospitalization. To assess the influence of a sample's acquisition time, we calculated the draw index, a numeric score between 0 and 1 calculated as the ratio of the blood draw

day over the duration of hospitalization. Accordingly, samples were assigned a hospitalization quarter ($0 \leq$ Quarter 1 $< 0{\cdot}25 \leq$ Quarter 2 $< 0{\cdot}50 \leq$ Quarter 3 $< 0{\cdot}75 \leq$ Quarter 4).

Bacterial Isolates

Bacterial isolates from 63 patients were recovered from blood culture, synovial fluid, or abscesses. Single colonies were selected and sub-cultured. *S. aureus* was confirmed by nuclease PCR and isolates were tested for methicillin resistance by mecA PCR. SCCmec typing of MRSA isolates was performed by classifying the ccr and mec complexes[181]. agr locus typing was performed and genetic relatedness was determined by repetitive-element, sequence-based PCR (rep-PCR). Gene encoding of toxins was detected by traditional PCR[181,182] (Figure 14, Table 9).

Batch Correction

To prevent batch effect between training and test sets, principal variance component analysis (PVCA) was conducted using JMP Genomics (SAS Institute, Cary, NC) to identify sources of batch effect. Cohort number accounted for 58% of the variability observed (Figure 15A) and scatter plot visualization (Figure 15B) with ellipsoids (Figure 15C) revealed strong segregation of samples based on cohort. The batch correction algorithm CombatR[183] was used on the cohort variable to reduce its contribution to the global variance. The batch effect from the cohort was reduced to approximately 0% (Figure 15D, Figure 15E and Figure 15F).

Microarray procedures

PBMC study - Affymetrix cRNA preparation

From 2–5 micrograms of total RNA, double-stranded cDNA was generated as a template for single-round *in vitro* transcription with biotin-labeled nucleotides using the Affymetrix cDNA Synthesis and In Vitro Transcription kits (Affymetrix Inc., Santa Clara, CA). Biotinylated cRNA targets were then purified (Sample Cleanup Module, Affymetrix) and hybridized to the Affymetrix HG-U133A and B GeneChip arrays (Affymetrix Inc., Santa Clara, CA) according to the manufacturer's standard protocols.

PBMC study - Affymetrix gene chips

Arrays were scanned using a laser confocal scanner (Agilent). Global gene expression analysis was carried out using the Affymetrix HG-U133A and U133B GeneChips. The HG-U133 set contains 44,760 probe sets representing. 39000 transcripts derived from 33,000 human genes. Raw signal intensity values were normalized to the mean intensity of all measurements per gene chip and scaled to a target intensity value of 500 using the MAS 5.0 global scaling method to adjust for possible chip-to-chip variations in hybridization intensities (GeneChip Operating System version 1.0). Data were imported into GeneSpring software (version 7.3.1, Agilent) to perform the gene expression analyses, statistical testing, hierarchical clustering, and classification of samples.

PBMC study - cRNA preparation for Illlumina Hu6v2 beadchips

Double-stranded cDNA was obtained from 200 ng of total RNA and after *in vitro* transcription underwent amplification and labeling steps according to the manufacturer's instructions. 1.5 mg of amplified biotin-labeled cRNA was hybridized to the Illumina Sentrix Hu6 BeadChips according to the sample labeling procedure recommended by Illumina. (Ambion, Inc, Austin, TX).

PBMC study - Illumina Hu6v2 Beadchips

The Sentrix Hu6 BeadChips consist of 50mer oligonucleotide probes attached to 3-mm beads within microwells on the surface of the glass slide representing 48,687 probes. Slides were scanned on Illumina BeadStation 500 and Beadstudio software was used to assess fluorescent hybridization signals.

Whole blood study – cRNA preparation for Illumina Hu6v3 beadchips

Total RNA was isolated from the whole blood lysed in Tempus tubes using the MagMAXTM-96 Blood RNA Isolation Kit (Applied Biosystems, Foster City, CA) according to the manufacturer's instructions. Following extraction, an Agilent 2100 Bioanalyzer (Agilent, Palo Alto, CA) was used to measure RNA Integrity Numbers (RIN), and samples with RIN values greater than seven were retained for further processing. RNA concentration was measured using a Nanodrop 1000 (Nanodrop

Technologies, Wilmington, DE). Following RNA extraction and quality control analysis, Globin mRNA was depleted from total RNA using the GLOBINclearTM-Human 96-well format kit (Ambion, Austin, TX). This was followed by another round of RIN and concentration determinations for quality control purposes. 250 ng of RNA from all samples passing quality control were amplified and labeled using the Illumina TotalPrep-96 RNA amplification kit (Ambion, Austin, TX). 750 ng of amplified labeled RNA were hybridized overnight to Illumina HT12 V3 beadchips (Illumina, San Diego, CA). Following hybridization, each chip was washed, blocked, stained, and scanned on an Illumina BeadStation 500 following the manufacturer's protocols. The dataset described in this manuscript is deposited in the NCBI Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo, GEO Series accession number GSE30119).

In vitro antigen-presenting cell study – cRNA preparation for Illumina Hu6v3 beadchips

Total RNA was isolated from cell lysates using the RNeasy Mini-Kit (#74104, Qiagen) according to the manufacturer's instructions. Following extraction, an Agilent 2100 Bioanalyzer (Agilent, Palo Alto, CA) was used to measure RNA Integrity Numbers (RIN) for each sample. All samples with RIN values greater than seven were retained for further processing. RNA concentration was measured using a Nanodrop 1000 (Nanodrop Technologies, Wilmington, DE). 250 ng of RNA from all samples passing quality control were amplified and labeled using the Illumina TotalPrep-96 RNA amplification kit (Ambion, Austin, TX). 750 ng of amplified labeled RNA were hybridized overnight to Illumina HT12 V3 beadchips (Illumina, San Diego, CA). Following hybridization, each

chip was washed, blocked, stained, and scanned on an Illumina BeadStation 500 following the manufacturer's protocols.

Microarray data analysis

PBMC study

Using the Genespring^{TM} software program (Agilent), the expression value for each transcript per individual subject's sample was normalized to the median expression value of that transcript in samples from healthy controls. Class comparison analyses were performed on probe sets present in at least 75% of samples in each group (quality control (QC) probes). Non-parametric statistical testing (Wilcoxon-Mann-Whitney U-test; $p<0.01$ for class comparisons; $p<0.05$ for modular analyses with no multiple test corrections) was used to rank genes based on their ability to discriminate among pre-specified groups of patients. Final lists of significantly changed genes used in class comparisons were filtered to include only those transcripts that showed a 1.25-fold or greater fold change in expression level relative to the control group. Hierarchical clustering was applied to order genes according to expression levels. The list of the top-ranked genes from the *S. aureus* biosignature was created by ranking genes with the highest fold change difference and the most significantly different genes (by p value) between healthy controls and patients with infection.

For the cross microarray platform validation part of the analysis we performed a two-step procedure. First, we examined the genes that defined the *S. aureus* biosignature

obtained on the test and training set of subjects analyzed on the Affymetrix platform in a new set of 18 subjects (validation set) and run on the lllumina microarray platform. Transcript sequences from RefSeq were used to perfectly match corresponding valid gene probe sets on each platform; Affymetrix probe set data encoded by Mage-ML files in XML format were matched by GenBank accession numbers and reference sequence transcripts to the Illumina manifest probe mapping file (http://www.switchtoi.com/probemapping.ilmn) allowing for mapping of the significant Affymetrix gene probes to their corresponding gene probes on the Illumina platform. In the second step of the analysis, the Illumina gene list was applied to an independent validation set of subjects using an unsupervised scheme that allowed clustering of samples based solely on intrinsic gene expression levels.

Modular analysis

Rationale

Traditional microarray analysis usually involves the comparison of two or more study groups, generating large lists of transcripts that are significantly differentially expressed between groups based on arbitrarily defined thresholds. This results in significant amount of noise and can affect data interpretation and biomarker discovery[184,185]. The modular approach aims at identifying sets of coordinately expressed transcripts across hundreds of condition, and treating each of these sets as a functional unit of transcription, thereby reducing the dimensionality of the data.

PBMC modular analysis framework development

A detailed account of this module-based data mining analysis strategy has been reported elsewhere[145]. In summary, a reference dataset of PBMC samples spanning 8 immunogical conditions including was selected. This unsupervised method identified 28 transcriptional modules formed by genes coordinately expressed across multiple disease data sets, thus allowing functional interpretation of the microarray data into biologically useful information, using these modular subunits rather than considering all transcripts.

Whole blood modular analysis framework development

This analysis strategy has been described elsewhere[145]. A set of 62 transcriptional modules derived from 410 whole blood gene expression profiles was applied to the whole blood dataset described herein. Modules were annotated with Ingenuity Pathway Analysis (IPA) (Ingenuity Systems, Redwood City, CA), Pubmed, iHOP, and Novartis Gene Atlas (http://biogps.gnf.org) databases. Module transcript content and annotations are available online (http://www.biir.net/public_wikis/module_annotation/V2_Trial_8_Modules).

Antigen-presenting cell modular analysis framework development

The module construction process is summarized in Figure 29. Six independent datasets (Table 19) generated on the Illumina v3 platform were used as input and combined into a single 353 samples dataset. Each dataset's batch-corrected expression

data was processed and clustered independently. First, all probes that were not detected in any of the 353 samples (Illumina detection p-value > 0.01) were removed, leaving 22,916 probes present at least once. All raw signals were then scaled such that all values less than 10 were set to 10. Then, each sample was normalized to the median of its own donor's cRPMI expression (control median). For each probe, if the difference between the sample and its control median was at least 100 and the ratio of (sample/control median) is at least 1.5, then the signal was set to 1 (over-expressed). If the difference between the sample and its control median was at least 100 and the ratio of (sample/control median) is less than or equal to 0.67, then the signal was set to -1 (under-expressed). Otherwise, the signal was set to 0 (no change). The dataset was then summarized so that for each probe, each stimulus group was represented by a single score. The group scored a 1 if all samples in the group were 1. It scored a -1 if all samples in the group were -1. Otherwise, the group scored a 0. Each dataset was then clustered. Probes were clustered together if and only if the pattern of -1/0/1 across all stimulus groups was a perfect match.

Taking the six sets of clusters as input, we constructed a weighted co-cluster graph, a probe by probe matrix where the value of each cell (the weight, between 0 and 6) was set to the number of times $probe_i$ and $probe_j$ were found in the same cluster. The goal was to extract sets of probes that are most frequently clustered together, proceeding from the most stringent requirements to the least. To accomplish this, we employed an iterative algorithm. To begin, the maximum clique threshold was initialized to the number of input cluster sets (6) and a minimum seed size was chosen (10). The outer loop begun by creating an unweighted graph through application of the maximum clique threshold to the

weighted co-cluster graph such that a probe pair, or edge, was represented in the unweighted graph if and only if the corresponding weight in the co-cluster graph equaled or exceeded this threshold.

The inner loop started. The maximum clique, which is the largest set of probes such that all pairs of probes in the set are connected in the unweighted graph, was isolated. If the size of the probe set is smaller than the minimum seed size, the inner loop ended, the threshold was reduced by one, and the outer loop restarted. Otherwise, the probe set was at least as large as the minimum seed size and forms a module. It was removed from both graphs and named in accordance with the iterations in which it was found (i.e. a module extracted in the first iteration of the outer loop and the second iteration of the inner loop is designated M1.2). The inner loop then begun again with the reduced graphs. Each clique identified formed a module. Once all cliques with size greater or equal to 10 were extracted, the algorithm ended.

Module annotations / interpretation

Modules were annotated with Ingenuity Pathway Analysis (IPA) (Ingenuity Systems, Redwood City, CA), Literature Lab (Acumenta, Boston, MA), Pubmed and iHOP databases. Annotated modules contained multiple genes with similar function, cellular localization, or membership in known biologic pathways. Furthermore, the Novartis Gene Atlas (http://biogps.gnf.org) was queried for modules containing genes with expression limited to particular cell populations. Those modules without known shared characteristics between member genes were not annotated.

Flow Cytometry

PBMCs were isolated by density gradient centrifugation using Ficoll-hypaque technique from blood samples (3–8 mL) collected in ACD tubes. One million PBMCs were washed and incubated with conjugated antibodies for 15 minutes at room temperature, in dark conditions according to multi-color staining panels described below. Cells were washed with phosphate buffer and then fixed with 2% paraformaldehyde. Samples were run on a BD LSRII flow cytometer (BD BioScience, San Jose, CA); 50,000 events were acquired with the BD FACSDivaTM software according to the lymphocyte gate and analyzed using FlowJoTM software (Tree Star, Inc). Multicolor staining panels included: B cell panel: CD19 (ECD, Beckman-Coulter), CD20 (Pe-Cy5, BD Pharmingen), CD24 (PE, BD Pharmingen), CD27 (APC, BD Pharmingen), CD38 (Pe-Cy7, BD Biosciences) and IgD (FITC, SouthernBiotech). T cell panel: CD3 (Alexa700, BD Pharmingen), CD4 (Pacific Blue, Invitrogen), CD8 (APC-Cy7, BD Pharmingen), CD45RA (ECD, Beckman- Coulter), CD62L (Pe-Cy5, BD Pharmingen), CCR7 (Pe-Cy7, BD Pharmingen). Monocyte panel: CD14 (Pacific Blue, BD Pharmingen), CD16 (APC, Invitrogen), CD40 (PE, BD Pharmingen), CD86 (FITC, BD Pharmingen), HLA-DR (APC-Cy7, BD Pharmingen) and CD62L (ECD, Beckman-Coulter). Isotype controls: IgG1 (FITC, BD Biosciences), IgG1 (PE, BD Biosciences), IgG1 (ECD, Beckman-Coulter), IgG1 (Pe-Cy5, BD Biosciences), IgG1 (Pe-Cy7, BD Biosciences), IgG1 (APC, BD Biosciences), IgG1 (Alexa 700, BD Biosciences), IgG1 (APC-Cy7, BD Biosciences), IgG1 (Pacific Blue, Invitrogen). Cell populations were

analyzed based on the following markers: B cells: CD19; Naive B cells: CD19+/CD20+, IgD+, CD27-; Memory B cells: CD19+/CD20+/IgD+&-/CD27+; Plasma cells: CD19+/CD20-/CD27+&++/CD38++; Transitional B cells: CD19+/CD20+/CD24++/CD38++; Pre-germinal center B cells: CD19+/CD20+/CD27+/CD38++; T cells: Naive T cells: CD3+/CD4+ or CD8+/CD45RA+/CCR7+; Central memory T cells: CD3+/CD4+ or CD8+/CD45RA-/CCR7+; Effector memory T cells: CD3+/CD4+ or CD8+/CD45RA-/CCR7-; CD8+ terminally differentiated T cells: CD3+/CD8+/CD45RA+/CCR7-.

Monocyte-derived dendritic cell activation *in vitro*

Monocyte-derived dendritic cell cultures

Monocytes were obtained from frozen fraction 5 from healthy donor apheresis. Cells were thawed for one minute in a 37°C water bath and resuspended in 1x phosphate buffered saline (PBS). Cells were spun at 350g for 7 minutes, washed with 1x PBS, counted, washed again, and resuspended in PBS / 2% Fetal Bovine Serum (FBS) / 1mM EDTA at $5x10^7$ cells/mL. Monocytes were then enriched using the EasySep Human Monocyte Enrichment Without CD16 Depletion Kit (#19058, StemCell Technologies) according to the manufacturer's protocol. Once enriched, cells were counted and resuspended in serum-free CellGenix DC medium (#2005, CellGenix, Germany) / 1% Penicillin/Streptomycin at $1x10^6$ cells/mL. GM-CSF was added for all DC subsets at 100 ng/mL. For IL-4 DC, recombinant IL-4 was added at 50 ng/mL and cells were fed a full

dose of GM-CSF and IL-4 at day 2 and day 4 of a 6-day culture. For IFN-alpha DC, IFN-alpha was added at 500U/mL, and cells were fed a full dose of GM-CSF and IFN-alpha at day 1 of a 3-day culture. For IL-15 DC, IL-15 was added at 100ng/mL, and cells were fed a full dose of GM-CSF and IL-15 at day 1 and day 3 of a 5-day culture. Cell suspensions were injected into 72mL sterile culture bags (#2PF-0072-AC, AFC, Gaithersburg, MD) with a 50mL syringe and bags were closed with injection ports. Feeding of the cells during the culture was done with a 1mL syringe through the injection ports. Cells were cultured in an incubator at 37°C, 5% CO_2.

Monocyte-derived dendritic cell activation

At the end of cell cultures, cells were collected, washed in 1x PBS and resuspended in complete RPMI at $1x10^6$ cells/mL. For the 353 samples reference dataset, all stimulations were conducted in 1.5mL polypropylene Eppendorf tubes in a total volume of either 1mL ($1x10^6$ cells) or 500uL ($5x10^5$ cells). All stimuli used are summarized in Table 20. At the end of the 6h stimulation period, cells were spun at 350g for 7 minutes, washed with 1x PBS, spun again and lysed in 600uL ($1x10^6$ cells) or 350uL ($5x10^5$) of RLT buffer (#79216, Qiagen). Cell lysates were stored at -80°C until extraction.

CHAPTER 3: *EX VIVO* ANALYSIS OF PBMC FROM PATIENTS WITH INVASIVE *STAPHYLOCOCCUS AUREUS* INFECTIONS

Introduction

When I joined the Ramilo laboratory in March 2007, the study on PBMC transcriptional profiles of patients with *S. aureus* infection had been recently initiated by Dr. Monica Ardura, who led the work presented in this chapter. The first goal of this study was to identify a conserved signature of *S. aureus* infection in peripheral blood mononuclear cells that could discriminate patients from healthy controls. The second goal was to conduct multicolor flow cytometry to compare the distribution of blood leukocyte populations in patients and healthy controls. Finally, we wanted to correlate PBMC transcriptional profiles to cell counts obtained by flow cytometry, to predict cell subpopulations contribution to the global fingerprint.

Patient Characteristics

Over a period of four years, samples from 53 previously healthy patients hospitalized with invasive *S. aureus* infections and 24 healthy control subjects were analyzed. Patients were chosen representing the clinical spectrum of acute severe *S. aureus* disease including bacteremia, osteomyelitis, suppurative arthritis, pyomyositis, and pneumonia with empyema. Patients with a diagnosis of staphylococcal toxic shock

syndrome or polymicrobial infections were excluded. Patient demographic data, clinical characteristics, analysis group, and microarray platform are summarized in Table 1. There were no statistical differences between the *S. aureus*-infected patients and their respective healthy controls with regards to age, sex, or race in the training and test sets (Table 2). Patients were enrolled only after a bacteriologic diagnosis was established; the median time from patient hospitalization to procurement of study blood sample was 4 days [IQ range 3–8 days]. Viral direct fluorescent antibody testing or culture of the nasopharynx was obtained on 68 subjects (88%, 46 patients, 22 controls) and did not reveal the presence of a concomitant viral infection.

Patients with culture-proven invasive *S. aureus* infections were divided into three groups for analysis: training, test, and validation sets. The training set of subjects composed of 20 children with invasive *S. aureus* infections (median age 7.5 years; 11 methicillin-resistant *S. aureus*, MRSA, and 9 methicillin-susceptible *S. aureus*, MSSA) and 10 healthy controls (median age 6 years) matched for age, sex, and race, was initially analyzed to identify the gene expression profile in patient PBMCs. As expected, statistical differences in laboratory parameters were observed between groups. *S. aureus* patients displayed higher total peripheral white blood cell count and percent neutrophil count, but lower percent lymphocyte count and hematocrit values in patients (Table 2). The test set included an independent group of 22 patients with *S. aureus* infection (median age 7 years; 8 MSSA, 14 MRSA) and 10 healthy controls (median age 6 years) and was used to validate the gene expression profile in PBMCs from *S. aureus*-infected patients. As in the training set, there were differences in laboratory values between patients and healthy controls (Table 2). A third independent group of 25 subjects was

included to validate our initial findings using (1) a second microarray platform (Illumina) and (2) flow cytometry. This validation set was comprised of 11 patients with *S. aureus* infection (median age 8 years; 5 MSSA, 6 MRSA) and 13 healthy controls (median age 9 years). PBMCs from 23 subjects were evaluated by flow cytometry to determine the relative abundance of different immune cell populations. Simultaneous flow cytometry evaluation and gene expression analysis was conducted with the same PBMC samples in 18 (9 with *S. aureus* infection and 9 healthy controls, matched for age, sex, race) of these 23 subjects. There were no statistical differences in antimicrobial therapy, or demographic and laboratory data between the training, test, and validation sets (Table 3).

Results

The PBMC signature from S. aureus patients is distinct from that of healthy controls

Statistical group comparison (Mann-Whitney, $p<0.01$) was applied to the list of transcripts present in the training set, revealing 3,168 genes differentially expressed between *S. aureus*-infected patients and healthy controls. The list was further filtered to include those transcripts with a 1.25-fold or greater change in expression level relative to the healthy control group, yielding a total of 3,067 transcripts. Hierarchical clustering was applied in order to visualize the transcriptional pattern (Figure 4A). For purposes of validation, the 3,067 transcripts list comprising the gene expression profile in PBMC of *S. aureus*-infected patients was then evaluated in an independent test set of 22 new patients. The samples were organized into a condition tree utilizing the 3,067 genes

obtained from the training set, and correctly classified 31 of 32 samples as either healthy or *S. aureus* infection based on the gene expression patterns (Figure 4B).

Genes represented in the expression profile in PBMCs from *S. aureus*-infected patients were then ranked according to differences in both fold-change and significance in gene expression levels ($p<0.05$) compared with healthy controls. The top 50 genes that were significantly over-expressed in patients with *S. aureus* infection versus healthy controls are shown in Table 4. Over-expressed genes included those with microbicidal functions (lactotransferrin, alpha-defensins 1 and 4, bactericidal/permeability-increasing protein), involved in coagulation (thrombomodulin), hemoglobin synthesis (hemoglobin D and G), pro-inflammatory and immune-related genes related to cellular growth, proliferation, and apoptosis pathways (ADM, ARG1, CLU, EGR1, IL8, HBEGF, ITGA2B, MMP9) and those involved in cell to cell signaling such as CEACAM6 and CEACAM8.

Module-level analysis reveals over-expression of innate and under-expression of adaptive immune response transcripts

To better characterize the biological significance of the gene expression profiles seen in the PBMCs of patients with *S. aureus* infections, gene expression levels between patients and healthy controls were mapped using a modular analysis framework that we previously described[145]. A key to the functional interpretation of each PBMC transcriptional module is detailed in Appendix A.

Gene expression levels were compared between patients and healthy controls on a module-by-module basis. The percentage of genes with a significant change (Mann-Whitney $p<0.05$) within each module are graphically displayed on a module map, with over-expressed genes represented in red and under-expressed genes in blue (Figure 5A, upper panel). Following the approach described previously with the class comparisons analysis, module analysis was applied initially to the training set of patients. Patients with *S. aureus* infection demonstrated significant over-expression of genes in modules related to innate immunity including myeloid lineage (M1.5, M2.6), neutrophil (M2.2), and inflammation (M3.2, M3.3) modules and under-expression of genes regulating adaptive immunity such as B cell module M1.3, cytotoxic cell module M2.1, and T cell specific module M2.8.

Significantly over-expressed genes included those in modules M1.5 and M2.6 (Myeloid lineage), containing transcripts related to myeloid cells, which are involved in bacterial pathogen recognition such as TLR2 and CD14, IL10 signaling (CD32, BLVRA) and leukocyte extravasation signaling (CTNNA1, NCF2, PECAM1, ITGB2), also in M2.6 genes related to the inflammatory response (calgranulin B, NFKB inhibitor, TNF superfamily members, and metalloproteinase inhibitors); in M2.2 (Neutrophils) genes encoding innate molecules including LTF, DEFA 1 and 4, BPI, CEACAM 8; in modules M3.2 and M3.3 (Inflammation I and II) genes involved in inflammatory and endothelial cell processes. Significant genes found over-expressed in M3.2 (Inflammation II) included those involved in coagulation (THBD), endovascular inflammation, TLR signaling (IRAK 3, Ly96), and transcriptional regulation (zinc finger proteins) while M3.3 included genes encoding antigens to scavenger receptor proteins (CD36),

coagulation (factor V), and lysosomal functions (LAMP2). Patients also displayed over-expression of module M2.3 (Erythrocytes), containing transcripts related to hemoglobin (hemoglobin alpha and gamma, erythrocyte membrane protein, erythroid factors), and module M3.5 (Undetermined) including hemoglobin alpha and gamma proteins.

Conversely, there was significant under-expression of transcripts related to the adaptive immune response including B cells (M1.3, with genes encoding cell surface molecules CD19, CD22, CD72, CD79 or involved in immunoglobulin production), T cells (M2.8, including CD6, CD96, ITK), cytotoxic cells (M2.1, including KLR subfamilies, Granulysin, Granzyme B) and genes encoding TNF family members. Module M1.8 (Undetermined) included under-expression of factors involved in DNA replication and transcription (zinc finger protein genes) and cytokines (IL16); significant under-expressed genes in M3.8 (Undetermined) also included multiple zinc finger proteins and TNF receptor-associated factors (TRAF5).

The transcript list forming the signature of *S. aureus* infection was then analyzed in the test set of patients and the initial findings were validated by module analysis. Both the gene probes and the percentage of over or under-expressed genes per module in patients with *S. aureus* infection in the test set matched those observed in the training set of patients, confirming the qualitative and quantitative consistency of the PBMC transcriptional signature in patients with *S. aureus* infection (Figure 5A, lower panel).

Confirming the robustness of the signature across microarray platforms

Twenty-five additional subjects were enrolled to further validate the PBMC gene expression profile of *S. aureus*-infected patients by confirming its reproducibility across microarray platforms. This validation set included 11 patients with invasive *S. aureus* infections and 14 age and sex-matched healthy controls. Transcript sequences from RefSeq were used to match corresponding valid gene probe sets on each platform. This allowed mapping of the 3,067 significant genes comprising the gene expression profile of PBMC of *S. aureus*-infected patients based on the Affymetrix gene probes to their corresponding 1,521 gene probes on the Illumina platform. This 1,521 gene list was then applied to the independent validation set of 9 patients and 9 controls using an unsupervised scheme that allowed clustering of samples based solely on intrinsic gene expression levels (Figure 5B). Despite technical differences between the two platforms, all 9 patients with *S. aureus* infection clustered together (red horizontal bar) based on similarities in PBMC gene expression patterns alone.

Module level analyses of the 1,521 gene probes on the Illumina platform demonstrated similarities between the Affymetrix and the Illumina data. As illustrated in Figure 5B, 16 of 19 modules yielded concordant results in both the Illumina and Affymetrix platforms. Only modules M1.4, M3.5, and M3.6 did not show significant differences in Illumina and may be a result of representation of less gene probes in those modules. Correlation analyses of the average normalized values among the differentially expressed genes for each given module between both platforms was statistically significant (Spearman $R=0.76$, $p=0.002$). Thus, both unsupervised hierarchical clustering and modular analyses confirmed the robustness of the gene expression profile in PBMC of *S. aureus*-infected

patients across both the Affymetrix and Illumina BeadChip microarray platforms in distinct groups of subjects.

Decreased number of central memory CD4+ and CD8+ T cells in patients with S. aureus infections

To understand how the changes in blood transcriptional profiles in patients with *S. aureus* infection relate to changes in the distribution of circulating leukocyte subpopulations, detailed flow cytometry phenotyping was conducted both in PBMCs from a subset of patients with *S. aureus* (n=11) and appropriate age-matched healthy controls (n=13).

As illustrated in Figure 6, there were no significant differences in the total number of B cells and T cells between patients with *S. aureus* infection and healthy controls ($p<0.05$) with the number of subjects considered (n=11). Given the significant under-expression of genes related to these cell populations observed in the modular analysis (M1.3, M2.1, M2.8), further detailed flow analysis was performed in each lymphocyte compartment. Characterization of B cell subpopulations (Figure 7) revealed no significant differences in the absolute number of naive (CD19+/CD20+, IgD+, CD27-), memory (CD19+/ CD20+/IgD+&-/CD27+), and plasma cells (CD19+/CD20-/ CD27+&++/CD38++) between patients and healthy controls (Figure 4). Transitional B cells (CD19+/ CD20+/CD24++/CD38++) were increased in patients compared with healthy controls ($p =0.04$); there was also a trend ($p=0.06$) toward increased numbers of

pre-germinal B cells (CD19+/CD20+/CD27+/CD38++) in patients versus control subjects.

We then analyzed both CD4+ and CD8+ T cell compartments, focusing on subpopulations of naive and memory cells based on the expression of CD45RA, CD62L, and CCR7 as previously described[186]. As illustrated in Figure 8, there were no differences in absolute numbers of naive CD4+ T cells and effector memory CD4+ T cells between patients and healthy controls. However, the absolute number of central memory CD4+ T cell was significantly reduced in *S. aureus* patients compared with controls ($p = 0.003$). The same observation was made for the CD8+ T cell compartment, with no differences between *S. aureus* patients and controls in the number of naive, effector memory, and terminally differentiated CD8+ T cells (Figure 9), but a significant decrease in central memory CD8+ T cells in *S. aureus* patients ($p = 0.005$). Thus, PBMCs of patients with invasive *S. aureus* infections demonstrate a significant reduction in the numbers of circulating central memory T cells in both CD4+ and CD8+ compartments.

Antigen-presenting cells expansion in patients with acute S. aureus infection

As shown in Figure 6, there was a significant increase in the median total number of monocytes in PBMCs of *S. aureus* patients when compared to healthy controls ($p = 0.002$). Detailed characterization of CD14+ subpopulations displayed a significant increase in absolute number of circulating monocytes expressing activation markers such as CD86, CD40, HLA-DR, and CD62L homing lymphocyte molecule in patients with *S. aureus* infection compared with controls (Figure 10A). Reanalysis of the CD14+

subpopulations was performed after removing outlier patients (patients 908, 952, and 960) in each subpopulation; *S. aureus* patients still demonstrated statistically significant increased numbers of CD14+ HLADR+ ($p = 0.023$), CD14+ CD40+ ($p = 0.0076$), CD14+ CD62L+ ($p = 0.0062$), and CD14+86+ ($p = 0.0062$) when compared to healthy controls.

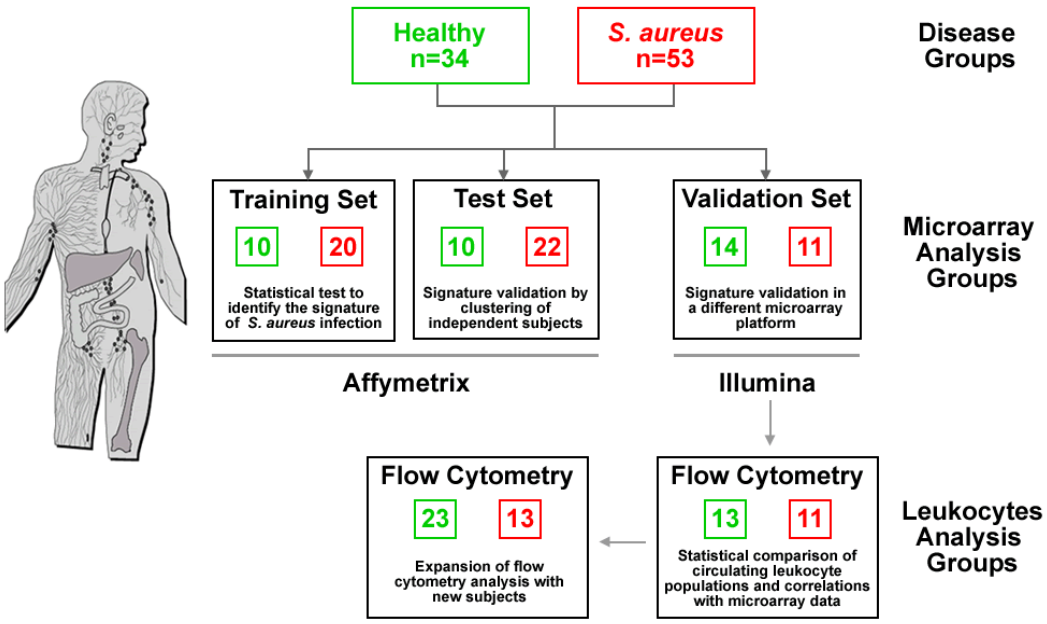
Based on the CD16 expression, circulating monocytes can be divided into functionally distinct subpopulations: CD14+CD16- monocytes and CD14+16+ monocytes. Patients with *S. aureus* infection showed a significant expansion of both CD14+16- ($p = 0.0021$) and CD14+16+ ($p= 0.0175$) monocytes compared with healthy controls (Figure 10B). These differences remained significant even when outliers were removed (CD14+16-, $p = 0.0076$; CD14+16-, $p = 0.0324$).

We subsequently expanded the flow cytometry study to 23 patients with *S. aureus* infection and compared them to 13 healthy controls. With these larger numbers, we additionally identified significant increases in circulating B cells, pDC and mDC (Figure 11A). All subpopulations of B cells but plasma cells were significantly increased (Figure 11B). Thus, PBMCs of patients with invasive *S. aureus* infections demonstrate a significant increase in the numbers of circulating antigen-presenting cells, including monocytes, pDC, mDC and B cells. It is interesting to notice that the microarray signature revealed a down-regulation of the B cell module. A potential explanation could be that the large expansion in circulating myeloid cells may cover the more subtle expansion in circulating B cells at the RNA level.

Gene expression levels correlate with specific immune cell populations

To further characterize the relationship between the gene expression profiles in PBMCs of patients with *S. aureus* infections and the changes of the different immune cell populations, analyses were conducted on a group of patients (n=9) and healthy controls (n=9) in whom simultaneous gene expression and flow cytometry analyses were performed. To this end, the per module PBMC average gene expression obtained from patients with *S. aureus* infection relative to the healthy controls, were run on the Illumina platform and correlated with the absolute number of immune cells in each subject as measured by flow cytometry (Table 5). The most significant statistical correlations observed in patients with *S. aureus* infections were between the total number of CD14+ monocytes and CD14+ monocyte subpopulations and several gene expression modules (Figure 12). Absolute number of CD14+ monocytes significantly positively correlated with myeloid cell (M1.5 and M2.6), neutrophil (M2.2), and inflammation II (M3.3) modules (Spearman $R>0.8$, $p<0.006$). Total CD14+ monocyte cell numbers inversely correlated with the modules encoding genes for ribosomal proteins (M2.4) and T cells (M2.8). The number of CD14+16- cells positively correlated with M2.6 (Myeloid lineage); CD14+16+ cells correlated with myeloid lineage modules (M1.5 and M2.6), neutrophils (M2.2), and inflammation (M3.3). The absolute number of CD14+CD162 and CD14+CD16+ cells inversely correlated with the ribosomal proteins (M2.4) and T cells (M2.8) modules. Further correlations performed in the monocyte compartment of CD14+CD62L+, CD14+ HLA-DR+, CD14+CD40+, and CD14+CD86+ revealed significant positive correlations between these subpopulations and the myeloid lineage module M2.6 and inverse correlations with ribosomal proteins (M2.6) and T cells (M2.8) modules (Figure 13). Gene expression levels in modules related to neutrophils (M2.2)

and inflammation (M3.3) positively correlated with CD14+CD62L+, CD14+ HLADR+, CD14+CD40+, and CD14+CD86+ cell numbers. With respect to the T cell compartment, the number of total CD4+ T cells inversely correlated with myeloid lineage (M1.5 and M2.6) and inflammation (M3.3) modules ($p = 0.0361$) (Table 5). Numbers of naive CD4+ T cells positively correlated with the modules related to ribosomal proteins (M2.4) and T cells (M3.8), while inversely correlating with the M2.6 (Myeloid) and inflammation (M3.2 and M3.3) modules. The number of CD4+ T central memory cells (TCM) correlated inversely with myeloid lineage (M1.5 and M2.6), neutrophil (M2.2), and inflammation (M3.3) modules, but positively with the ribosomal protein (M2.4), T cell (M2.8), and interferon (M3.1) modules. CD8+ T central memory cell numbers correlated inversely with myeloid modules (M1.5 and M2.6) and inflammation (M3.3) modules; conversely, CD8+ T effector memory cells correlated positively with cytotoxic cell module (M2.1) and undetermined (M1.8 and M3.7) modules. There were no significant correlations between modular gene expression levels and CD4+ T effector memory, CD8+ naive, and CD8+ terminally differentiated T cells. There were no significant correlations between the total number of CD19+ B cells or B cell subpopulations such as naive B cells, memory B cells, transitional B cells, pre-germinal center B cells, and plasma cells B cell and modular gene expression patterns (Table 6).



Panel 1: Design for the study on transcriptional profile of PBMC from patients with *S. aureus* infection

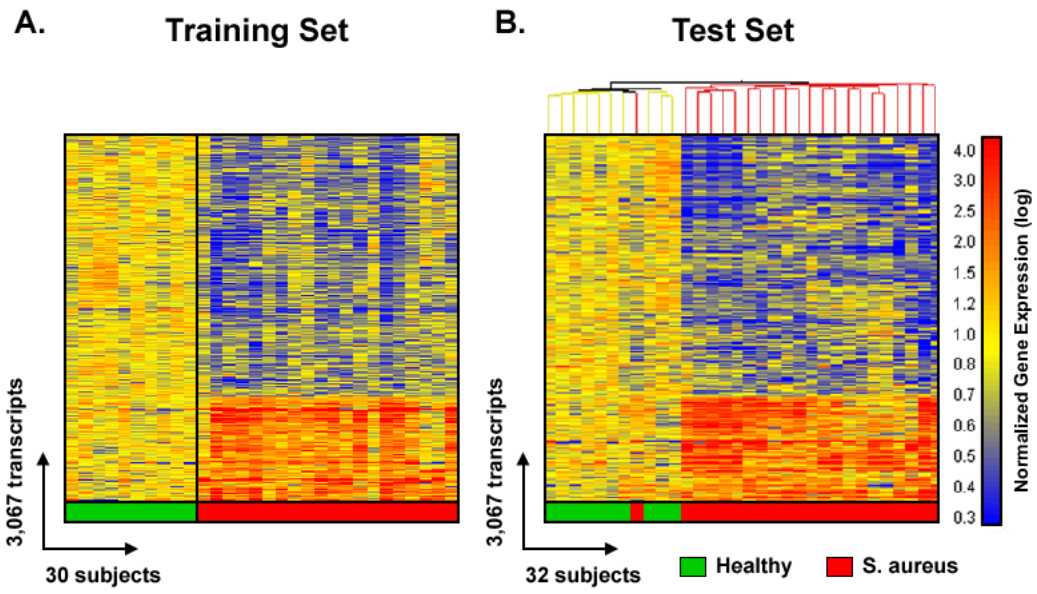


Figure 4: Gene expression biosignature in PBMCs from *S. aureus* patients and healthy controls.

A. Statistical group comparisons between 10 healthy subjects and 20 patients with acute *S. aureus* infections yielded 3,067 genes expressed at statistically different levels (Mann-Whitney $p<0.01$ and 1.25 fold change) between the two groups. Significant genes were organized by hierarchical clustering to reveal differential expression, each row representing a single gene and each column an individual subject. Transformed expression levels are indicated by color scale: red representing relatively high and blue relatively low gene expression compared to the median expression for each gene across all patients compared to healthy controls. **B.** The same 3,067 genes list was used to perform a condition tree on 22 new subjects with *S. aureus* infections and correctly grouped 21 of the 22 patients based solely on gene expression.

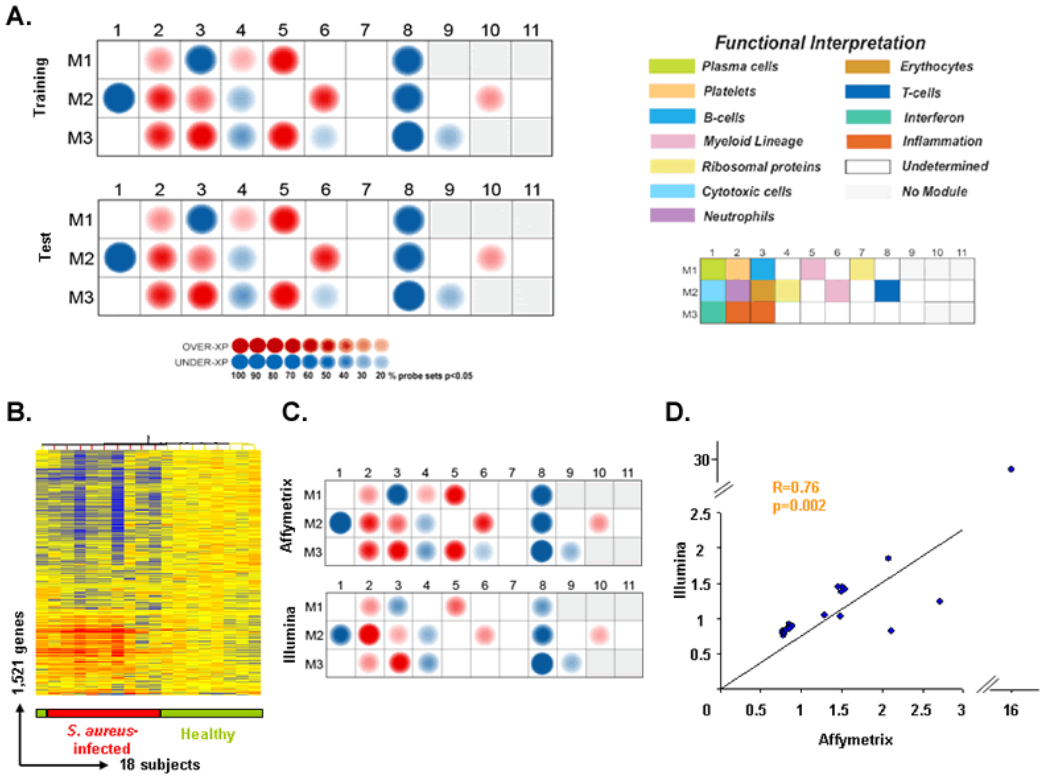


Figure 5: Module analysis identifies a specific fingerprint in the PBMCs of *S. aureus*-infected patients.

A. Gene expression levels were compared between patients with *S. aureus* infections and healthy controls on a module-by-module basis. Colored spot represent the percentage (color intensity) of significantly over-expressed (red) or under-expressed (blue) transcripts ($p<0.05$, Mann Whitney) within a module in patients with *S. aureus* infections; blank modules demonstrate no significant differences between groups ($p>0.05$). Information is displayed on a grid, with the coordinates corresponding to one of 28 modules with the key (upper right panel) representing the functional interpretation of modules. Nineteen modules are shown to be significantly different between healthy

subjects and patients with *S. aureus* infection in the training set (A., upper module map). The same gene list applied to the independent test set of patients reveals a similar fingerprint (A, lower module map). **B.** Transcript sequences (RefSeq) were used to map the corresponding 3,067 significant gene probe sets comprising the gene expression profile of PBMC in patients with *S. aureus* infection on the Affymetrix platform to their corresponding 1,521 gene probes on the Illumina platform (B, far left panel) and tested in new *S. aureus* patients (n = 9) and controls (n = 9). **C.** Comparison of module analysis in Affymetrix (top map) and Illumina (bottom map). **D.** Correlation analyses (Spearman) between the average gene expression levels per module in each platform were performed.

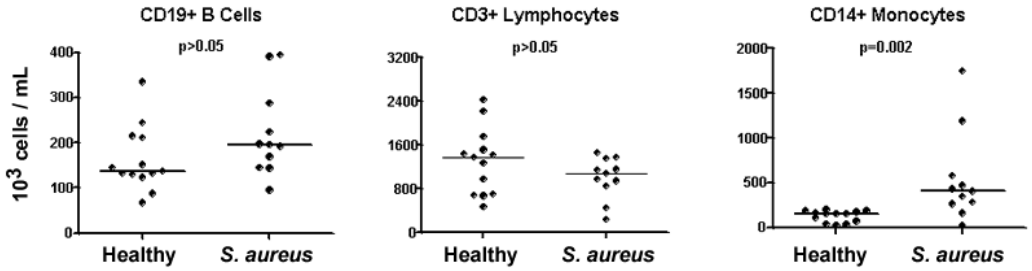


Figure 6: Peripheral blood monocytes are significantly expanded in patients with invasive *S. aureus* infections.

PBMCs obtained from age-matched healthy donors (n = 13) and patients with *S. aureus* infection (n = 11) were analyzed by flow cytometry for the expression of CD19 (left panel), CD3 (middle), and CD14 (right) markers. Results are expressed as absolute number of cells per mL of blood. Bars represent median values. Mann-Whitney test was applied for statistical analysis.

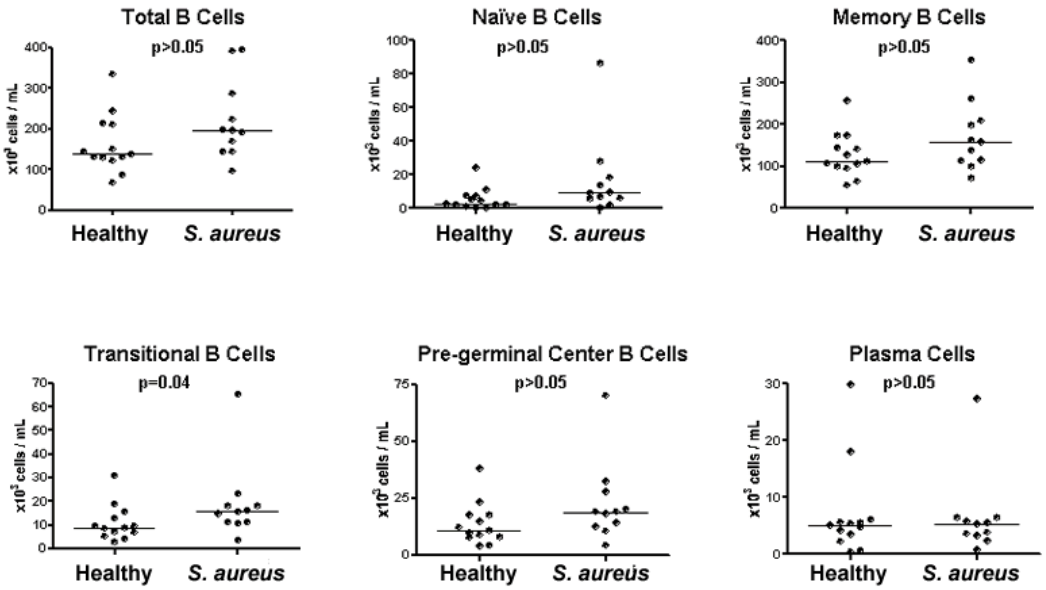


Figure 7: Analysis of B cell subpopulations in patients with *S. aureus* infection

PBMCs obtained from age-matched healthy donors (n = 13) and patients with *S. aureus* infection (n = 11) were stained by multicolor panel staining according to the expression of specific B cell markers. Results are expressed as absolute number of cells per mL of blood. Bars represent median values. Mann-Whitney test was applied for statistical analysis.

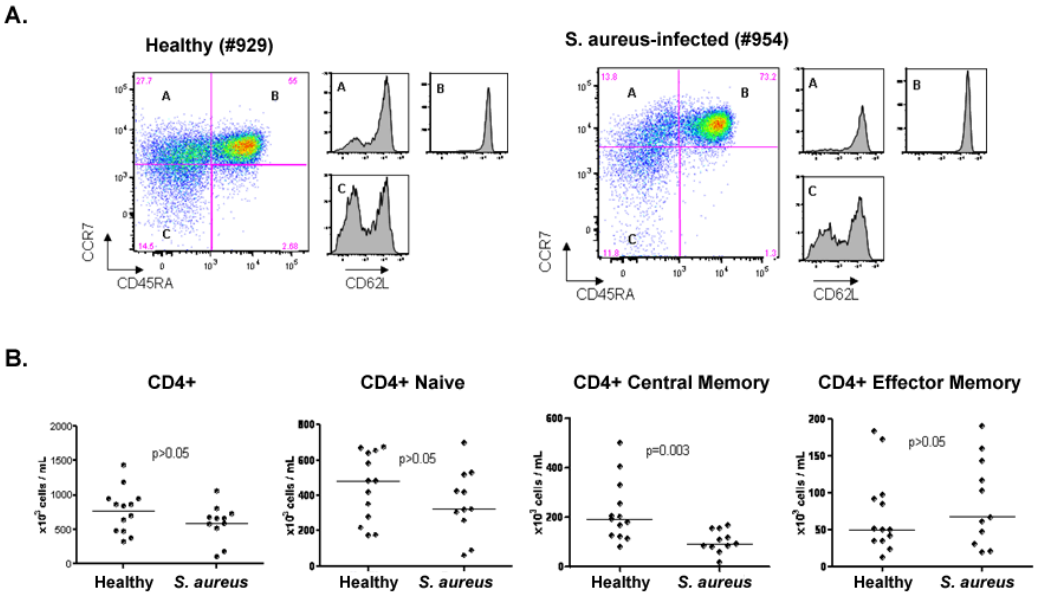


Figure 8: Decreased central memory CD4+ T cells in patients with *S. aureus* infection

PBMCs obtained from age-matched healthy controls (n = 13) and patients with *S. aureus* infection (n = 11) were stained with CD3, CD4, CCR7, CD45RA, and CD62L antibodies and analyzed by flow cytometry. CD4 T cell subsets are labeled as (A) central memory T cells (TCM), (B) naive T cells, and (C) effector memory T cells (TEM). A. Flow cytometry plots of CD4 T cell subsets in a representative *S. aureus* patient and healthy control. CD62L expression is shown for each CD4 T cell subset. B. Dot-plots displaying the numbers of CD4 T cells and CD4 T cell subsets expressed as absolute number of cells per mL of blood. Horizontal lines represent median values. Statistical analysis was performed using a Mann-Whitney test.

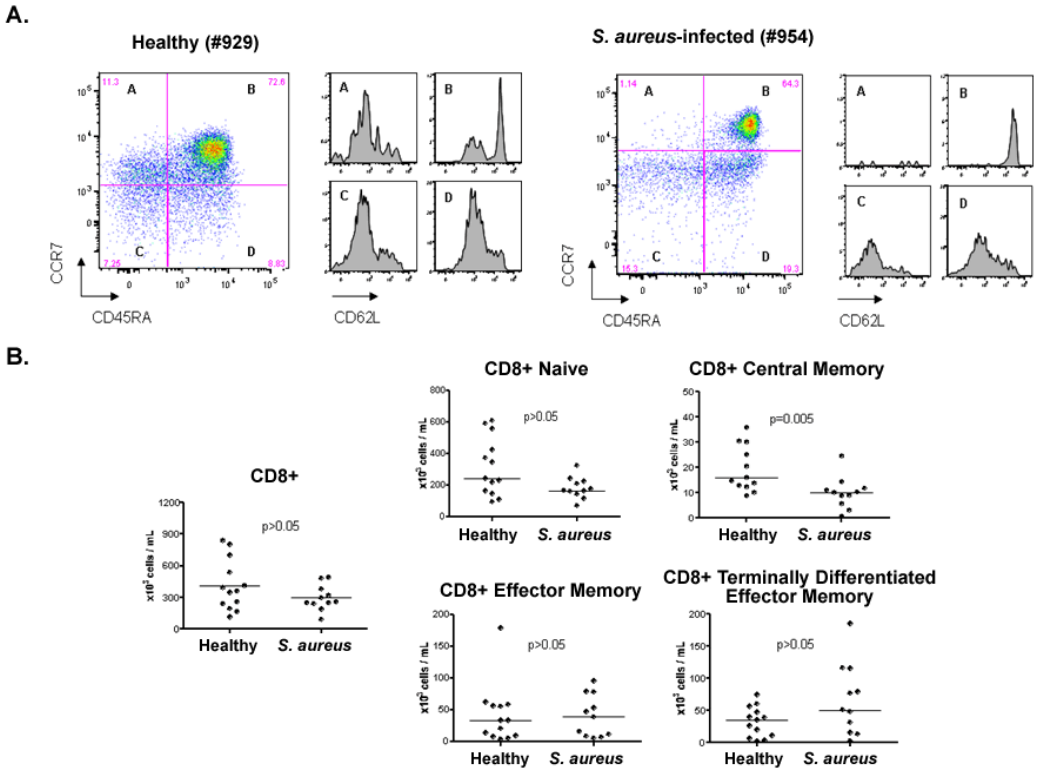


Figure 9: Decreased central memory CD8+ T cells in patients with *S. aureus* infection.

PBMCs obtained from patients with *S. aureus* infection (n = 11) and age-matched healthy controls (n = 13) were stained with CD3, CD8, CCR7, CD45RA, and CD62L antibodies and analyzed by flow cytometry. CD 8 T cell subsets are labeled as: (A) central memory T cells (TCM); (B) naıve T cells, (C) effector memory T cells (TEM), and (D) terminally differentiated effector T cells (TEM). **A.** Flow cytometry plots show CD8 T cell subsets in a representative *S. aureus* patient and healthy control. CD62L expression is shown for each CD8 T cell subset. **B.** Dot plots displaying the numbers of CD8 T cells and CD8 T cell subsets. Results are expressed as absolute number of cells per mL of blood.

Horizontal lines represent median values. Statistical analysis was performed using a Mann-Whitney test.

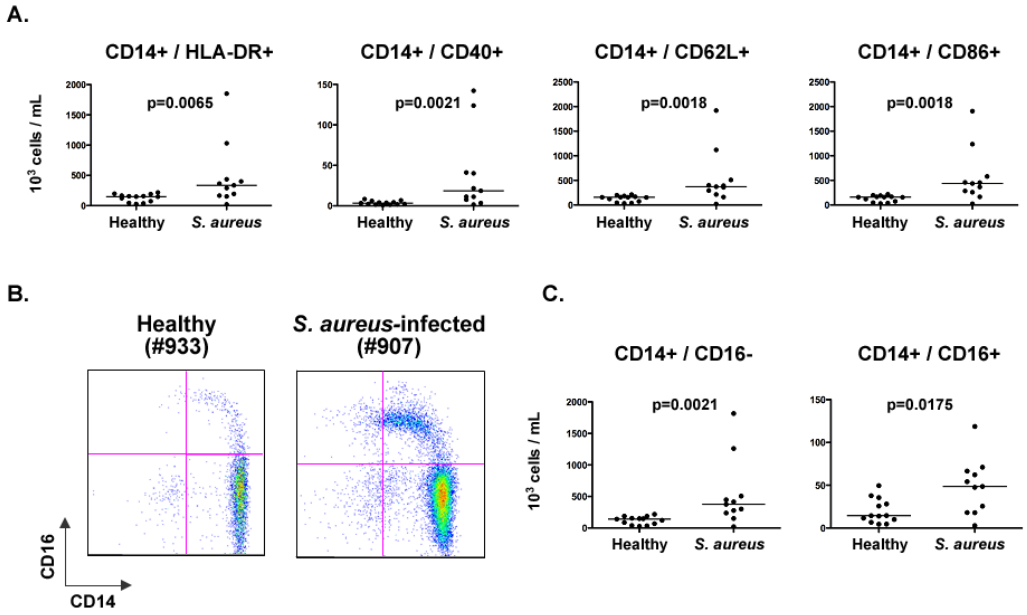


Figure 10: Increased numbers of activated CD16-monocytes and inflammatory CD16+ monocytes in *S. aureus* patients.

PBMCs obtained from patients with *S. aureus* infections (n=11) and age-matched healthy controls (n=13) were stained with HLA DR, CD40, CD62L, CD16, and CD86 antibodies and analyzed by flow cytometry. **A.** Graphs show the absolute monocyte cell numbers. Results are expressed as absolute number of cells per mL of blood. Horizontal lines represent median values. Statistical analysis was performed using a Mann-Whitney test. **B.** Cytometry plots showing PBMC gated on CD14+ cells, for a representative *S. aureus* patient (907) and healthy control (933). **C.** Dot plots displaying the absolute number of circulating CD14++ CD16- and CD14+ CD16+ monocytes in *S. aureus* patients versus healthy controls.

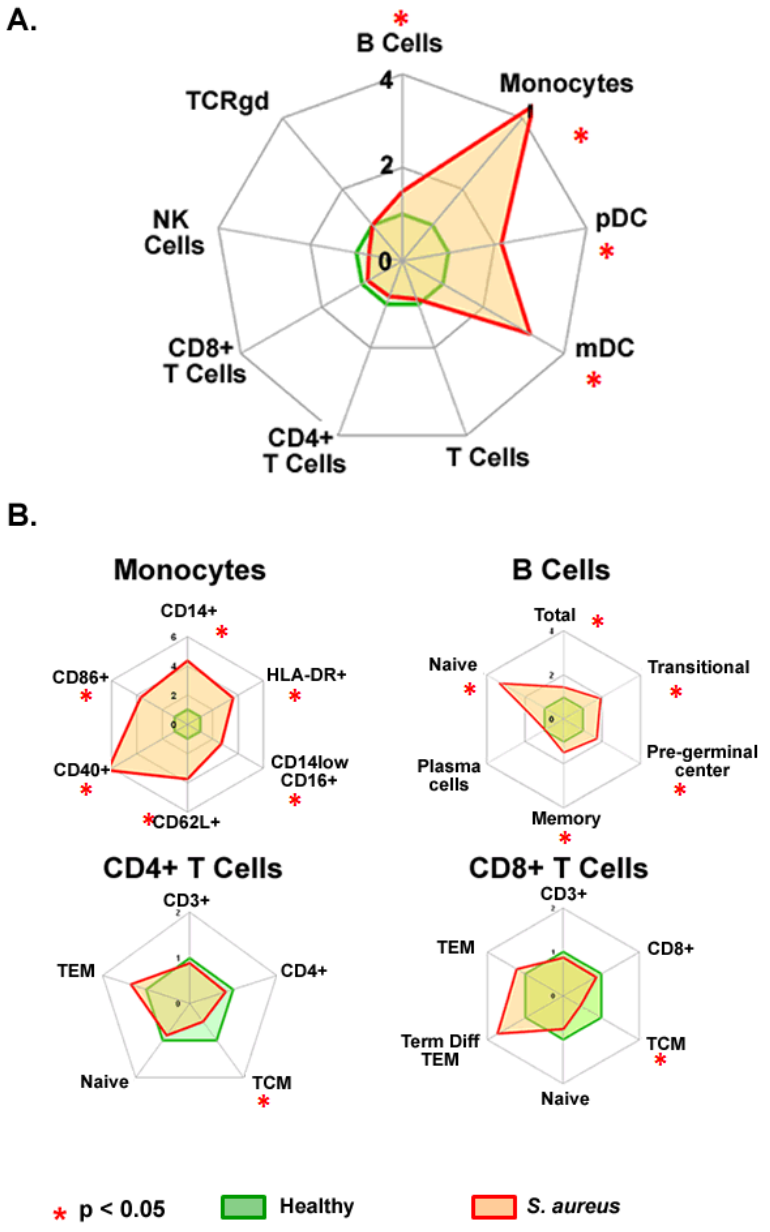


Figure 11: Increase circulating APC in patients with *S. aureus* infections

A. Radar chart displaying the average fold change in absolute cell population count in patients (red area) versus healthy controls (green area). Data are normalized to the

average of healthy controls. **B.** Radar charts displaying subpopulation fold change in patients for monocytes, B cells, CD4+ and CD8+ T cells. Red asterisks highlight significantly increased or decreased populations ($p<0.05$).

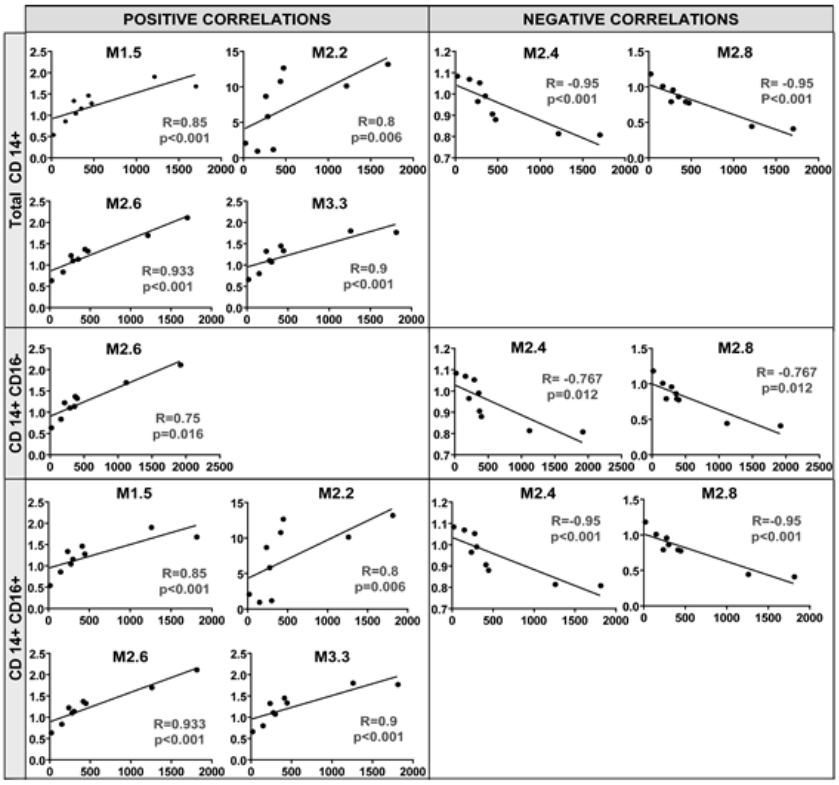


Figure 12: Significant correlations between monocyte subpopulation counts and specific modules.

Correlation analyses were performed between the significant modules comprising the gene expression profile of PBMCs from *S. aureus*-infected patients and monocyte populations. Graphs represent the correlation (Spearman) between the normalized average gene expression [log] significantly changed (Mann Whitney $p<0.05$) in the PBMCs of patients with *S. aureus* infection relative to the median gene expression of PBMCs in healthy controls in each significant module (y axis) and the corresponding absolute cell count by subpopulations per mL of blood in *S. aureus*-infected patients (x axis).

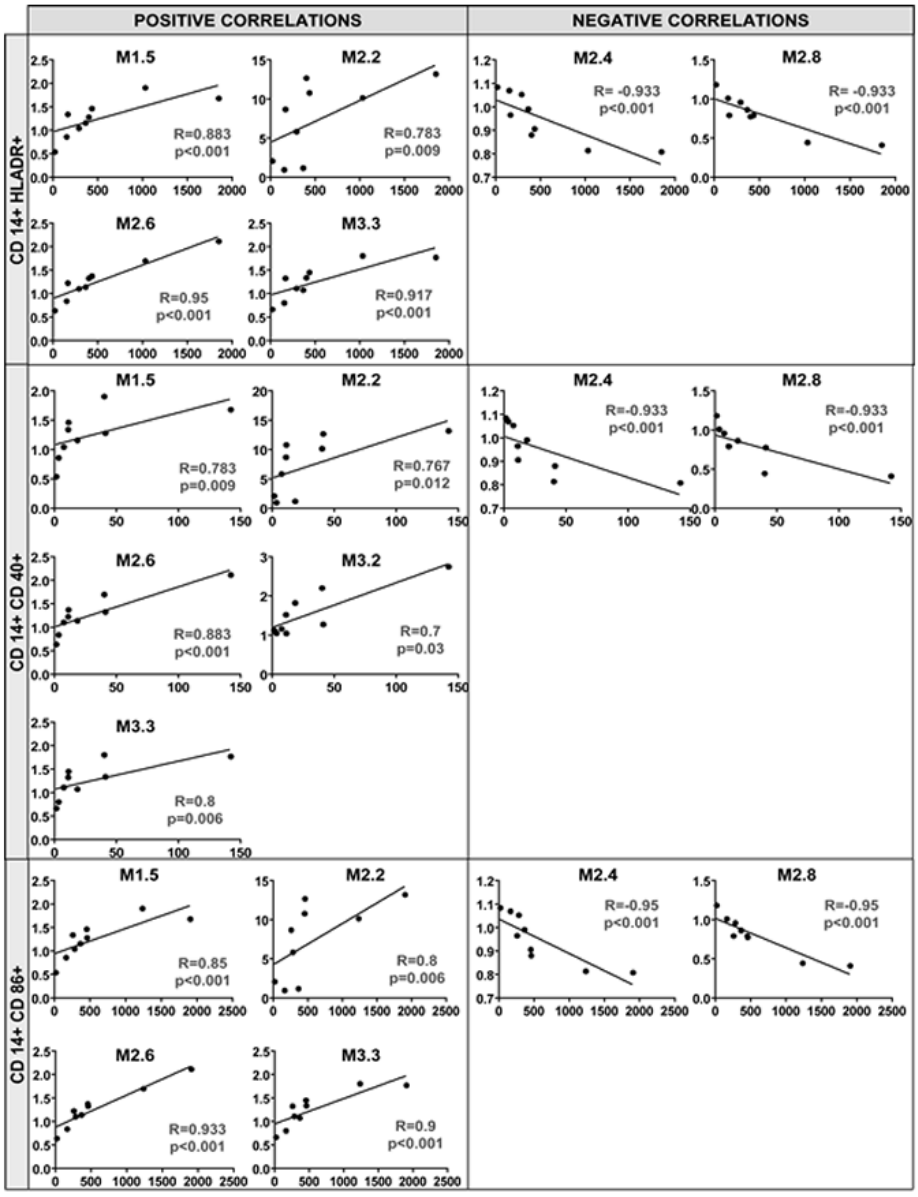


Figure 13: Significant correlations between activated CD14+ monocyte counts and specific modules.

Graphs represent correlations (Spearman) between average gene expression [log] significantly changed (Mann Whitney $p<0.05$) in the PBMCs of patients with *S. aureus* infection relative to the median gene expression of PBMCs in healthy controls in each significant module (y axis) and the corresponding activation marker-specific absolute cell count per mL of blood in *S. aureus*-infected patients (x axis).

Subject	Age (yrs)	Ethnicity	Gender	Condition	Pathogen	Analysis Set	Platform
3N	6	H	M	Control	Healthy	Training/Test	A
5	10	H	M	Osteomyelitis	MSSA	Training	A
7N	1.6	B	F	Control	Healthy	Training/Test	A
8N	10	B	M	Control	Healthy	Training/Test	A
11N	4	B	F	Control	Healthy	Training/Test	A
23N	7	H	F	Control	Healthy	Training/Test	A
24	3	B	M	Bacteremia, Osteomyelitis, Myositis	MRSA	Test	A
24N	3	W	M	Control	Healthy	Training/Test	A
30	15	B	M	Bacteremia	MRSA	Test	A
40	13	W	M	Bacteremia, Osteomyelitis	MSSA	Training	A
43	7	B	M	Bacteremia, Osteomyelitis, SArthritis, Pyomyositis, Emboli	MRSA	Test	A
62	2	W	M	Osteomyelitis, Pyomyositis	MRSA	Training	A
66	0.25	B	F	Pneumonia	MRSA	Test	A
67	7	W	F	Bacteremia, Osteomyelitis	MRSA	Training	A
88	0.92	H	M	Bacteremia, Osteomyelitis, Pneumonia, Emboli	MRSA	Test	A
90	0.67	B	M	Bacteremia, SArthritis	MSSA	Test	A
109	0.67	H	F	Bacteremia, SST Abscess	MRSA	Training	A
150	9	B	F	Bacteremia, Osteomyelitis, SArthritis, Myositis	MRSA	Test	A
179	12	W	M	Bacteremia, Endocarditis, Emboli	MSSA	Training	A
205	7	H	M	Bacteremia, Pneumonia, SST Abscess	MRSA	Test	A
208	10	W	F	Bacteremia, Osteomyelitis, CNS abscess, Pneumonia, Emboli	MRSA	Test	A
216	10	H	F	Bacteremia, Osteomyelitis	MRSA	Training	A
220	11	H	M	Bacteremia, Osteomyelitis	MSSA	Test	A
221	6	B	F	Bacteremia, Osteomyelitis	MRSA	Training	A
224	10	W	M	Bacteremia, Osteomyelitis	MSSA	Test	A
230	20	B	M	Bacteremia, Endocarditis, SST Abscess	MRSA	Test	A
241	0.92	B	F	Bacteremia, Osteomyelitis, Pneumonia	MRSA	Training	A
242	1.2	B	M	Bacteremia, Osteomyelitis, Pyomyositis	MRSA	Test	A
258	8	W	F	Bacteremia, Osteomyelitis, Cellulitis	MSSA	Training	A
262	13	H	M	Bacteremia, SST Abscess	MRSA	Training	A
264	13	B	M	Bacteremia, Osteomyelitis, SArthritis, SST Abscess, Myositis, Emboli	MSSA	Test	A
271	13	B	M	Osteomyelitis, pyomyositis	MSSA	Training	A
294	12	B	F	Control	Healthy	Training/Test	A
301	8	W	M	Control	Healthy	Training/Test	A
303	6	W	F	Control	Healthy	Training/Test	A
304	6	W	M	Control	Healthy	Training/Test	A
305	4	H	F	Bacteremia, Osteomyelitis, SArthritis	MSSA	Training	A
308	12	B	F	Bacteremia, Pneumonia, Pyomyositis, SST Ab	MSSA	Training	A
328	0.38	B	F	Bacteremia, Osteomyelitis, Pneumonia, SST Abscess	MSSA	Test	A
329	0.58	O	F	Bacteremia, Lymphadenitis	MRSA	Training	A
330	11	B	F	Bacteremia, Pneumonia	MRSA	Test	A
354	2	H	F	Bacteremia, Osteomyelitis	MRSA	Training	A
366	11	B	F	Osteomyelitis, SST Abscess	MRSA	Test	A
369	14	B	M	Bacteremia, SArthritis, SST Abscess, Emboli	MRSA	Test	A
372	14	W	M	Bacteremia, Osteomyelitis, Myositis	MRSA	Test	A
412	1.75	W	M	Bacteremia, SArthritis	MSSA	Test	A
418	8	H	F	Osteomyelitis	MSSA	Training	A
423	8	W	M	Bacteremia, Osteomyelitis, Pyomyositis	MSSA	Training	A
434	5	W	M	Osteomyelitis, Pyomyositis	MSSA	Test	A
440	0.5	B	M	Bacteremia, Osteomyelitis, Pneumonia, Emboli	MRSA	Training	A
450	0.83	B	M	Bacteremia, Osteomyelitis, SArthritis, Pyomyositis	MSSA	Test	A
451	4	H	M	Bacteremia, Osteomyelitis, SArthritis,	MRSA	Training	A

				Pyomyositis, Emboli			
894	8	W	F	Osteomyelitis, SST Abscess	MSSA	Validation, FACS	I
903	13	W	F	Control	Healthy	Validation, FACS	I
904	16	W	F	Control	Healthy	FACS	I
905	16	W	F	Control	Healthy	FACS	I
907	1.3	H	F	Bacteremia	MRSA	FACS	I
908	8	B	M	Bacteremia, Pleural effusion, Pyomyositis, SST Abscess, Emboli	MSSA	FACS	I
909	6	H	F	Bacteremia, Osteomyelitis, SArthritis, Pneumonia	MRSA	Validation, FACS	I
910	5	B	M	Osteomyelitis, SArthritis	MSSA	Validation, FACS	I
922	3	H	F	Control	Healthy	FACS	I
926	3	H	M	Control	Healthy	FACS	I
927	2	H	M	Control	Healthy	Validation, FACS	I
929	11	O	F	Control	Healthy	Validation, FACS	I
930	8	O	M	Control	Healthy	Validation	I
931	12	W	F	Control	Healthy	FACS	I
933	9	W	M	Control	Healthy	Validation, FACS	I
935	9	W	F	Control	Healthy	Validation, FACS	I
939	8	W	M	Control	Healthy	Validation, FACS	I
940	7	W	M	Control	Healthy	Validation, FACS	I
941	9	W	F	Control	Healthy	Validation, FACS	I
943	9	W	M	Osteomyelitis, SArthritis, Pyomyositis	MRSA	Validation, FACS	I
944	6	B	F	Bacteremia, Osteomyelitis, SArthritis, Pyomyositis	MRSA	Validation, FACS	I
949	11	W	M	Bacteremia, Osteomyelitis, SArthritis	MSSA	Validation	I
952	15	B	M	Bacteremia, Osteomyelitis, Pneumonia	MRSA	Validation, FACS	I
954	13	W	M	Bacteremia, Osteomyelitis, Pyomyositis	MSSA	Validation, FACS	I
960	3	H	M	Pneumonia	MRSA	Validation, FACS	I

H=Hispanic, W=White, B=Black, O=Other; M=Male, F=Female; SST Ab=Skin/soft tissue abscess, SArthritis=Suppurative arthritis, CNS=central nervous system; MSSA=methicillin-susceptible *Staphylococcus aureus*, MRSA=methicillin-resistant *Staphylococcus aureus*; Platform: A=Affymetrix U133 A&B; I=Illumina Sentrix Hu6 BeadChips

Table 1: PBMC study - Subjects characteristics.

	TRAINING SET			TEST SET		
Parameter	***Patients***	***Controls***	***p[a]***	***Patients***	***Controls***	***p[a]***
Age (years)	7.5 [2-11]	6 [3.5-9]	0.86	7 [1-12]	6 [3.5-9]	0.93
Race	5B,8H,6W,1O	4B,2H,4W	1	14B,3H,5W	4B,2H,4W	0.25
Gender	9M:11F	5M:5F	0.57	16M:6F	5M:5F	0.45
WBC (10^3/mm^3)	8.9 [7.5-17.4]	7.2 [5.2-8.2]	**0.03**	11 [7.7-16]	7.2 [5.2-8.2]	**0.009**
Neutrophils (%)	60 [48-73]	43 [27-50]	**0.007**	61 [40-66]	43 [27-50]	**0.022**
Lymphocytes (%)	24 [13-38]	47 [39-55]	**0.002**	24 [12.7-47]	47 [39-55]	**0.038**
Monocytes (%)	8 [7-11]	7 [6-10]	0.37	9 [6.3-12.5]	7 [6-10]	0.38
Hematocrit (%)	31.6 [29.8-36]	37.7 [34.9-40]	**0.007**	30 [27-33.9]	37.7 [34.9-40]	**0.002**
Platelets (10^3/mm^3)	382 [297-459]	310 [264-329]	0.07	353 [273-445]	310 [264-329]	0.373
CRP (mg/dL)	6.65 [2.1-16.5]	0.4 [0.4-0.7]	**<0.001**	7.4 [2.3-16.9]	0.4 [0.4-0.7]	**0.001**

Table 2: Demographic and laboratory characteristics of patients and controls in training and test sets.

S. aureus patients	Training Set	Test Set	Validation Set	p-value[a]
Age (years)	7.5 [2-11]	7 [1-12]	8 [5-11]	0.82
Race	5B,8H,6W,1O	14B,3H,5W	4B,3H,4W	0.08
Gender	9M:11F	16M:6F	7M:4F	1
WBC (thousand/mm^3)	8.9 [7.5-17.4]	11 [7.7-16]	8.5 [5.7-15.5]	0.52
Neutrophils (%)	60 [48-73]	61 [37-66]	55 [45-71]	0.73
Lymphocytes (%)	24 [13-38.3]	24 [12.7-47.5]	29 [19.5-44.5]	0.70
Monocytes (%)	8 [7-11]	9 [6.3-12.5]	7 [3.5-11.5]	0.58
Hematocrit (%)	31.6 [29.8-36.2]	30 [27.2-33.9]	32 [29-34.1]	0.08
Platelets (thousand/mm^3)	382 [297-459]	353 [273-445]	364 [313-473]	0.61
ESR (mm/hr)	70 [34-98]	46 [36-72]	72 [46-87]	0.28
CRP (mg/dL)	6.7 [2.1-16.5]	7.4 [2.4-16.9]	7.7 [3.7-21.3]	0.69

Median values [25-75% range]; B=Black, H=Hispanic, W= White, O=Other; M=Male, F=Female; CRP=C-reactive protein; [a] p values calculated within each individual subject set and respective to matched controls (Kruskal-Wallis)

Table 3: Demographic and laboratory characteristics of patients in training, test and validation sets.

Rank	Common Name	Description	Fold Change	Significance
1	LTF	Lactotransferrin	46.3	2.90E-05
2	CEACAM8	Carcinoembryonic antigen-related cell adhesion molecule 8	32.0	5.20E-05
3		Transcribed locus	32.9	2.90E-05
4	IL8	Interleukin 8	21.6	7.50E-05
5	HBG1	Hemoglobin, gamma A	21.5	7.05E-04
6	DEFA4	Defensin, alpha 4, corticostatin	20.6	7.50E-05
7	DEFA1	Defensin, alpha 1, myeloid-related sequence	17.0	4.30E-05
8	HBG2	Hemoglobin, gamma G	15.8	1.54E-03
9	LCN2	Lipocalin 2 (oncogene 24p3)	13.7	9.00E-05
10	HBD	Hemoglobin, delta	13.3	4.32E-04
11	ARG1	Arginase, liver	14.4	3.66E-04
12	BPI	Bactericidal/permeability-increasing protein	12.8	1.29E-04
13	HPR	Haptoglobin-related protein	12.2	1.55E-04
14	S100P	S100 calcium binding protein P	11.6	5.20E-05
15	THBD	Thrombomodulin	9.0	3.09E-04
16	ADM	Adrenomedullin	8.7	2.00E-05
17	CA1	Carbonic anhydrase I	8.6	4.32E-04
18	ERAF	Erythroid associated factor	7.9	1.55E-04
19	MSCP	Mitochondrial solute carrier protein	7.6	4.30E-05
20	MYL9	Myosin, light polypeptide 9, regulatory	7.5	5.10E-04
21	MMP9	Matrix metalloproteinase 9	7.0	5.80E-03
22	RETN	Resistin	6.9	1.08E-04
23	ANXA3	Annexin A3	6.5	7.05E-04
24	CEACAM6		5.9	2.40E-03
25	MGAM	Maltase-glucoamylase (alpha-glucosidase)	5.8	4.32E-04
26	PBEF1	Pre-B-cell colony enhancing factor 1	5.6	1.54E-03
27	CYP4F3	Cytochrome P450, family 4, subfamily F, polypeptide 3	5.5	4.32E-04
28	PBEF	Pre-B-cell colony enhancing factor 1	5.4	6.00E-04
29	ITGA2B	Integrin, alpha 2b	5.2	8.27E-04
30	HP	Haptoglobin	5.1	3.50E-05
31	SNCA	Synuclein, alpha (non A4 component of amyloid precursor)	4.9	1.60E-05
32	FCGR3A	Fc fragment of IgG, low affinity IIIb, receptor for (CD16)	4.9	2.90E-05
33	EGR1	Early growth response 1	4.4	3.20E-03
34	LOC199675	Hypothetical protein LOC199675	4.4	3.09E-04
35	SOCS3	Suppressor of Cytokine Signaling 3	4.3	1.30E-05
36	EREG	Epiregulin	4.2	2.07E-03
37	DTR	Heparin-binding EGF-like growth factor	4.1	2.40E-03
38	TCN1	Transcobalamin I	4.0	7.28E-03
39	MS4A3	Membrane-spanning 4-domains, subfamily A, member 3	4.0	2.78E-03
40	GLUL	Glutamate-ammonia ligase (glutamine synthase)	3.9	3.66E-04
41		FP15737	3.9	6.20E-05
42	FLJ31978	Hypothetical protein FLJ31978	3.9	2.61E-04
43	CLU	Clusterin	3.7	8.27E-04
44	MAD	MAX dimerization protein 1	3.7	1.60E-05
45	H1F0	H1 histone family, member 0	3.7	4.30E-05
46	FCGR1A	Fc fragment of IgG, high affinity Ia, receptor for (CD64)	3.7	2.90E-05
47	SOCS3	Suppressor of cytokine signaling 3	3.7	1.30E-05
48	MS4A4A	Membrane-spanning 4-domains, subfamily A, member 4	3.6	1.79E-03
49	PBEF	Pre-B-cell colony enhancing factor 1	3.6	1.55E-04
50	ORF1-FL49	Putative nuclear protein ORF1-FL49	3.6	1.55E-04

Table 4: Top-50 over-expressed genes in invasive *S. aureus* infections.

FACS Cell Marker	Module	Assigned Immune Function	Correlation (Spearman)	p-value
CD19	*none*			
CD3	*none*			
CD4	M1.5	Myeloid	-0.683	0.0361
	M2.6	Myeloid	-0.650	0.05
	M3.3	Inflammation II	-0.683	0.0361
CD8	M3.1	Interferon	0.65	
NK	*none*			
CD14	M1.3	B cells	-0.633	0.0583
	M1.5	Myeloid	0.85	<0.001
	M2.2	Neutrophils	0.8	0.006
	M2.4	Ribosomal	-0.95	<0.001
	M2.6	Myeloid	0.933	<0.001
	M2.8	T cells	-0.95	<0.001
	M3.3	Inflammation II	0.9	<0.001

Table 5: Correlation between gene expression levels and absolute number of immune cells in patients with *S. aureus* infections.

FACS Cell Marker	Module	Assigned Immune Function	Correlation (Spearman)	p-value
CD4, naïve	M2.4	Ribosomal	0.683	0.036
	M2.6	Myeloid	-0.65	0.05
	M2.8	T cells	0.683	0.036
	M3.2	Inflammation I	-0.65	0.05
	M3.3	Inflammation II	-0.7	0.03
CD4 TCM	M1.5	Myeloid	-0.783	0.009
	M2.2	Neutrophils	-0.667	0.043
	M2.4	Ribosomal	0.65	0.05
	M2.6	Myeloid	-0.717	0.025
	M2.8	T cells	0.65	0.05
	M3.1	Interferon	0.7	0.03
	M3.3	Inflammation II	-0.8	0.006
CD4 TEM			none	
CD8, naïve			none	
CD 8 TCM	M1.5	Myeloid	-0.767	0.012
	M2.6	Myeloid	-0.683	0.036
	M3.3	Inflammation II	-0.783	0.009
CD8 TEM	M1.8	Undetermined	0.65	0.05
	M2.1	Cytotoxic cells	0.667	0.043
	M3.7	Undertermined	0.683	0.036
CD8-terminally differentiated			none	
B cells, naïve			none	
Memory B			none	
Transitional B			none	
Pre-germinal B			none	
Plasma B			none	

FACS= fluorescence-activated cell sorting; TCM= central memory T cell; TEM = effector memory T cell

Table 6 – Correlations of B and T cell subsets with transcriptional modules

CHAPTER 4: *EX VIVO* ANALYSIS OF WHOLE BLOOD FROM PATIENTS WITH COMMUNITY-ACQUIRED *STAPHYLOCOCCUS AUREUS* INFECTIONS

Introduction

The analysis of PBMC from patients with community-acquired *S. aureus* infection helped us identify a disease fingerprint common to all patients. Furthermore, we were able to correlate transcriptional changes with changes in frequency of circulating leukocytes subpopulations, including monocytes. Additionally, we observed the decrease in central memory T cells, which may be relocating to the peripheral lymphoid tissues to activate effector T cell and/or B cell responses against *S. aureus*. This study has however several limitations. 1) As transcriptional profiling of PBMC requires immediate manipulation of fresh blood samples, it is not easily applicable to large-scale inter-institutional clinical studies. 2) Neutrophil and erythrocyte depletion from PBMC results in the loss of important information for the systemic characterization of these infections[187], as these cells play an essential role in the defense against blood borne pathogen. 3) As we focused on common transcriptional changes induced by *S. aureus*, we did not investigate heterogeneous elements of the signature, nor their clinical correlates.

To address these remaining questions, we conducted a new study in which we collected whole blood samples in Tempus™ tubes (Applied Biosystems, CA) from 99 pediatric patients with community-acquired *S. aureus* infection and 44 healthy controls. Again, the patients displayed a wide spectrum of clinical presentations, including skin and soft tissues abscesses without bacteremia, osteomyelitis, suppurative arthritis,

pneumonia and disseminated disease affecting multiple loci. After validating in whole blood the observations made in PBMC regarding the common transcriptional elements induced by *S. aureus*, we investigated the blood transcriptional correlates of clinical heterogeneity using two different approaches. First, patients were clustered based on transcriptional profiles without *a priori* knowledge of clinical diagnoses, and statistical analyses were carried on to identify clinical correlates. In a second phase of the analysis patients were grouped according to extent of bacterial dissemination, and clinical presentation and analyses were performed to identify transcriptional correlates. The design of this study is summarized in Panel 2. We measured transcriptional responses both qualitatively, using a modular analytical framework specifically developed for whole blood, and quantitatively, using the molecular distance to health (MDTH), and transcriptional score of global signature perturbation.

Results

S. aureus induces a distinct whole blood transcriptional signature

99 patients and 44 healthy controls were assigned to independent training and test sets. The training set included 40 patients with *S. aureus* infection and 22 healthy controls matched for age, sex and race (Table 10). Statistical group comparison yielded 1,422 differentially regulated transcripts. Hierarchical clustering of these transcripts grouped them according to similarities in gene expression patterns (Figure 16A). This signature was validated in the independent test set of 59 patients and 22 healthy controls (Figure 16B). Hierarchical clustering of the 1,422 transcripts from the training set grouped 52 out

of 59 patients from the test set together. The seven patients who clustered with controls presented were closer to recovery and discharge, as defined by the mean draw index (0·68 vs. 0·53 for all other patients).

S. aureus induces innate immunity and suppresses adaptive immunity transcripts

We used a preexisting analytical framework of 62 transcriptional modules that group together genes with shared expression pattern across independent blood transcriptional datasets[145]. Module maps were derived independently for the training (Figure 16C) and test sets (Figure 16D), using their respective healthy control group as reference. A module annotation legend is provided (Figure 16E**)**. Patients with acute *S. aureus* infection demonstrated significant over-expression of modules linked to the myeloid lineage (M3.2, M4.6, M4.13, M4.14 and M6.6), and inflammation (M4.2, M5.1 and M6.13) confirming our earlier findings in PBMC. Furthermore, patients displayed over-expression of modules linked to the coagulation cascade (M1.1), hematopoietic precursors (M3.3 and M5.3), and neutrophils (M5.15). Conversely, they demonstrated significant under-expression of modules linked to T cells (M4.1, M6.15 and M6.19), cytotoxicity/NK cells (M3.6 and M4.15), B cells (M4.10), and lymphoid lineage (M4.7 and M6.9). Transcript composition of these modules is summarized online (http://www.biir.net/public_wikis/module_annotation/V2_Trial_8_Modules) and in Appendix B. These findings were confirmed in the test set as shown by significant correlation of module expression between training and test sets (Figure 16F, $p<0·0001$, Spearman $r = 0·94$).

Heterogeneity in the blood signature

As the clinical presentations of individual patients were diverse (Table 11), we hypothesized that this diversity might be reflected in their blood transcriptional signature. Because the training and test sets yielded similar modular fingerprints (Figure 16F), they were merged into one dataset for subsequent analysis. To assess the magnitude of individual transcriptional perturbations, the molecular distance to health, or MDTH (a score measuring the global change of the transcriptional profile as compared to the median of healthy controls[131,133]), was calculated for each individual subject. Patients displayed a median MDTH significantly higher than that of healthy controls (Figure 17A, $p<0{\cdot}001$). Variability was observed from patient to patient, and 25 of them displayed MDTH within healthy range (3-259). These patients were flagged as transcriptionally quiescent (TQ) and separated from transcriptionally active patients (n=74). Unsupervised hierarchical clustering of the 10,972 transcripts expressed in at least one of the 143 subjects grouped these quiescent patients with healthy controls (Figure 17B).

To assess the qualitative heterogeneity of active patients, individual modular fingerprints were derived and k-means clustering was used to further identify modular expression patterns. Four major patient clusters (C1 to C4) were identified (Figure 17C, Figure 17D) using the "jump method"[188] (Figure 18). Group module expression was derived for each cluster (Figure 17D, Figure 19). Cluster C1 included 13 patients with a high mean MDTH (3,003) and strong over-expression of myeloid (M3.2, M4.6, M4.13, M6.6) and inflammation (M4.2, M5.1, M6.13) modules. Cluster C2 comprised 31 patients with a signature qualitatively similar to C1, but quantitatively dimmer as

supported by the lower mean MDTH (973). Cluster C3 regrouped 12 patients with high MDTH (2,395) and low inflammatory signature, but over-expression of modules linked to erythropoiesis (M2.3, M3.1, M4.4, M6.18) and hematopoiesis (M5.3). C1 and C3 both displayed significant down-regulation of transcripts linked to B and T cells (Figure 20). Gene level analysis (Figure 21A) combined with PANTHER[189] ontology ranking (http://www.pantherdb.org) supported this observation as heme biosynthesis (Figure 21B) and porphyrin metabolism (Figure 21C) were the most enriched pathways in C3 versus C1. The top 100 transcripts differentially expressed between C1 and C3 are displayed in Table 12 and Figure 21D. Finally, cluster C4 regrouped 18 patients with low mean MDTH (699). Interestingly, 11 of them displayed an IFN signature, although only one patient had concomitant CMV infection detected by viral culture.

The median values for different clinical findings and laboratory parameters were calculated for each cluster (Figure 17E, Table 13). Cluster C1 displayed a lower draw index, indicating that blood samples were collected at an earlier stage of hospitalization. Patients from C1 had a longer duration of hospitalization than other patients. Additionally, patients from C1 had significantly higher CRP, WBC, neutrophil and monocyte counts. Thus, these routine laboratory markers of inflammation corroborated the over-expression of myeloid and inflammatory modules.

The C3 erythropoietic signature overlapped with transcripts from CD71+ early erythroid precursors[190] (Figure 21E). The C3 patients showed decreased hemoglobin, hematocrit, and MCHC (Figure 21F, Table 13), suggesting an anemic state that might trigger erythropoiesis as indicated by the signature.

Finally, we analyzed how infection site, clinical presentation, and bacterial strain were distributed among clusters (Figure 17F, Table 14). Cluster C1 contained most patients with severe infection and disseminated disease, and a higher proportion of patients with pneumonia. Clusters C4 and TQ included most patients with skin and soft tissue abscesses, confirming the association between low MDTH and mild presentation.

Bacterial isolates

To determine whether heterogeneity in bacterial virulence factors was associated with clustering patterns, the isolates from 63 patients were characterized. The great majority (87%) of isolates tested were PVL-positive. Both clusters C1 and C3 displayed a higher percentage of MRSA isolates (100% and 91·7%) than the overall mean (67%), suggesting that MRSA might induce a stronger host response. No difference in other bacterial characteristics such as agr locus type or genetic relatedness was observed between the four clusters (Table 15).

Elements of the molecular signature correlate with clinical parameters

We then asked whether patient's molecular profiles correlated with clinical laboratory parameters commonly used to assess clinical evolution (Table 16). The MDTH positively correlated with neutrophil count, white blood cell count, C-reactive protein, band neutrophil count, red blood cell distribution width, and monocyte count. MDTH inversely correlated with relative lymphocyte count, draw index, red blood cell count, hemoglobin concentration, mean corpuscular hemoglobin concentration, and hematocrit.

Correlations were then assessed on a module-by-module basis (Figure 22A). Inflammatory modules positively correlated with neutrophil counts, CRP and WBC while modules linked to adaptive immunity negatively correlated with these parameters. Two modules annotated as hematopoiesis (M3.3) and cell cycle (M6.11) correlated with absolute band (immature neutrophils) count. Significant correlations between clinical nodes and molecular nodes are summarized as a network (Figure 22B).

Blood draw index, infection dissemination and clinical presentation influence the fingerprint

Clinical and molecular correlates can be linked by grouping patients based on prior knowledge of clinical features, and assessing molecular differences between these groups. To this end, supervised analysis was conducted to compare transcriptional signatures according to: 1) time in the course of the infection as defined by draw index quarters; 2) infection site; 3) types of invasive clinical presentations. Five myeloid lineage-related modules were differentially regulated from quarter to quarter (Figure 23A). The MDTH also displayed a decreasing trend (Figure 23B) that was paralleled by the decrease in CRP (Figure 23C).

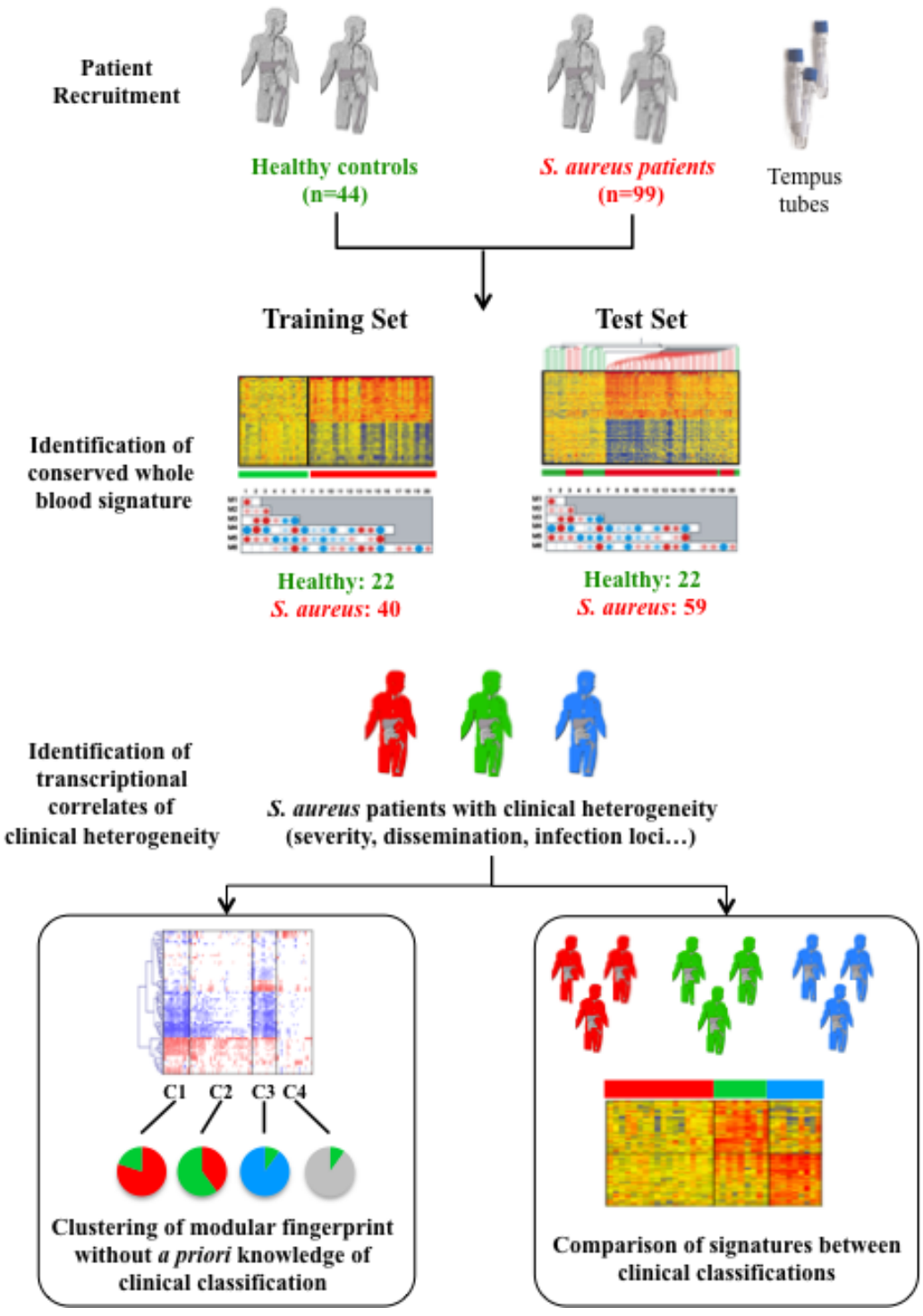
MDTH increased as the infection became more disseminated (Figure 23D), which was most evident during the first half of the hospitalization period (Figure 24A). Dissemination of infection was paralleled by decreased RBC, hemoglobin and hematocrit, and increased RDW, ESR, CRP and duration of hospitalization (Table 17). Finally, patients with pneumonia had a higher median MDTH, median WBC, neutrophil

and monocyte counts than patients with osteoarticular infections (Figure 23E, Figure 24B, Figure 25A, and Table 18).

Patients with osteoarticular infections display transcripts linked to blood coagulation

To assess transcriptional differences between patients with distinct clinical presentations, nine patients with osteoarticular infection were compared to nine patients with pneumonia and both groups were matched for MDTH (Figure 25B). 18 healthy controls (nine from each training and test sets) were used as reference. Module fingerprints identified over-expression of the coagulation cascade (M1.1) and platelet adhesion (M6.14) in osteoarticular infection but not in pneumonia (Figure 26A). From the 385 genes differently expressed (Figure 26B), PANTHER analysis for pathway enrichment (Figure 26C) identified blood coagulation as the most significant pathway in genes over-expressed in osteoarticular infection, and cholesterol biosynthesis in genes over-expressed in pneumonia. Thus, patients hospitalized with different clinical syndromes display distinct transcriptional profiles despite the fact that the infections were caused by the same bacterial pathogen.

We also determined whether specific *S. aureus* strains were associated with certain clinical manifestations, but we did not observe significant correlations between bacterial isolates (Table 15) and clinical presentation.



Panel 2: Whole blood study design

Whole blood was obtained for 99 patients and 44 healthy controls. Subjects were split into independent training and test set to identify and validate a global whole blood signature of *S. aureus* infection. Patients were subsequently recombined into a single group and transcriptional correlates of clinical heterogeneity were identified by two different methods: 1) Patient modular fingerprints were clustered without a priori knowledge of clinical classification and distribution of clinical observations were tested between transcriptional clusters. 2) Patients were first grouped according to clinical phenotype and statistical tests were conducted between these groups to identify modules and transcripts differentially regulated.

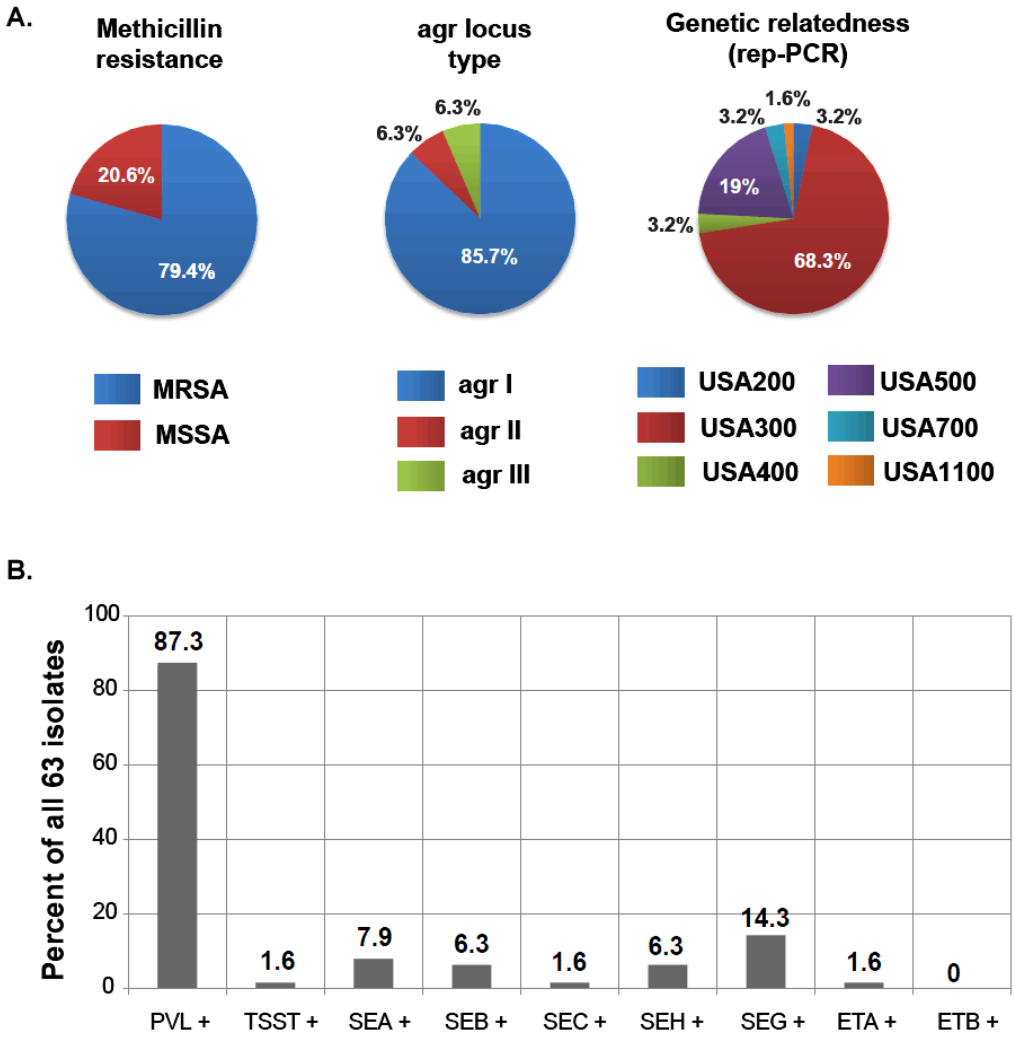


Figure 14: Characterization of the 63 cultured bacterial isolates.

A. Pie charts representing distribution of methicillin resistance, agr locus type and genetic relatedness for the 63 cultured bacterial isolates. **B.** Bar chart representing the percentage of cultured isolates positive for the measured toxins.

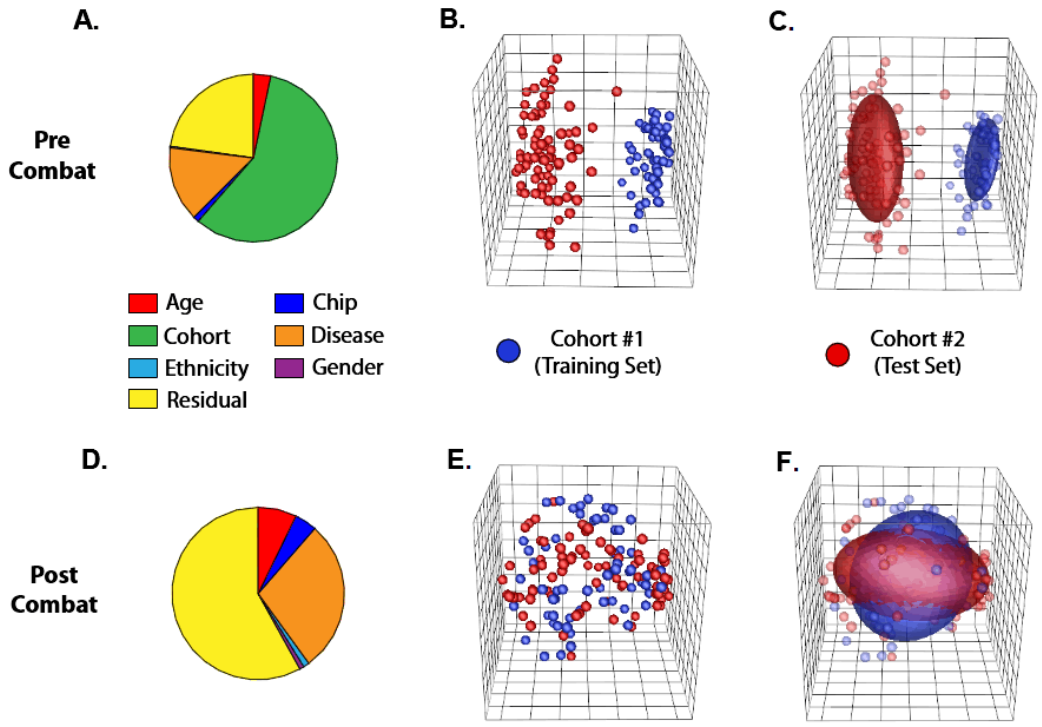


Figure 15: Batch correction for the two cohorts of patients used as training and test sets.

A. Pie chart of individual parameter's percent contribution to variance from principal variance component analysis (PVCA) before CombatR correction. **B.** Scatter plot representing the segregation of sample by cohort before CombatR correction. **C.** Elliptical fit of scatter plot data from B. **D.** Pie chart of individual parameter's percent contribution to variance from PVCA after CombatR correction for cohort. **E.** Scatter plot representing the segregation of sample by cohort after CombatR correction. **F.** Elliptical fit of scatter plot data from E.

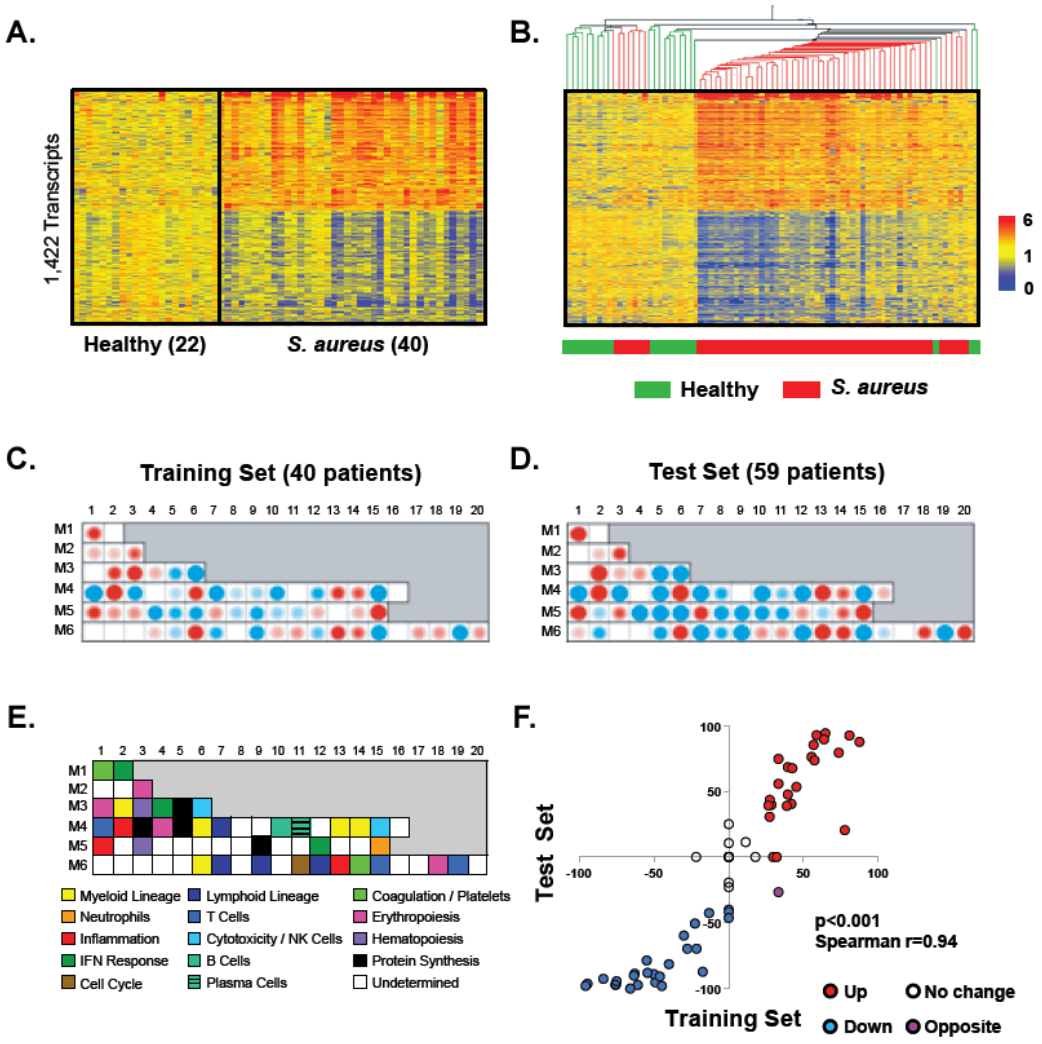


Figure 16: The *S. aureus* infection whole blood transcriptional signature is characterized by over-expression of myeloid lineage transcripts and under-expression of lymphoid lineage transcripts.

A. Statistical group comparison between 22 healthy subjects and 40 patients with acute *S. aureus* infection (non-parametric test, α=0·01, Benjamini-Hochberg multiple testing correction, 1·25 fold change) yielded 1,422 differentially expressed transcripts .

Transcripts were organized by hierarchical clustering (Spearman) according to similarities in expression profiles. Each row represents a transcript and each column an individual subject. Normalized log ratio levels are indicated by red (over-expressed) or blue (under-expressed), as compared to the median of healthy controls. **B.** The same 1,422 transcript list and hierarchical clustering were applied to an independent test set of 22 healthy controls and 59 patients with acute *S. aureus* infection. Sample hierarchical clustering (Spearman) was performed on the 1,422 transcript list in the test set. **C.** Average modular transcriptional fingerprrint for *S. aureus* patients as compared to healthy controls in the training set. **D.** Average modular transcriptional profile for *S. aureus* patients as compared to healthy controls in the test set. **E.** Module functional annotations legend. **F.** Scatter plot comparing module expression between training (x-axis) and test (y-axis) sets. Spearman correlation was applied.

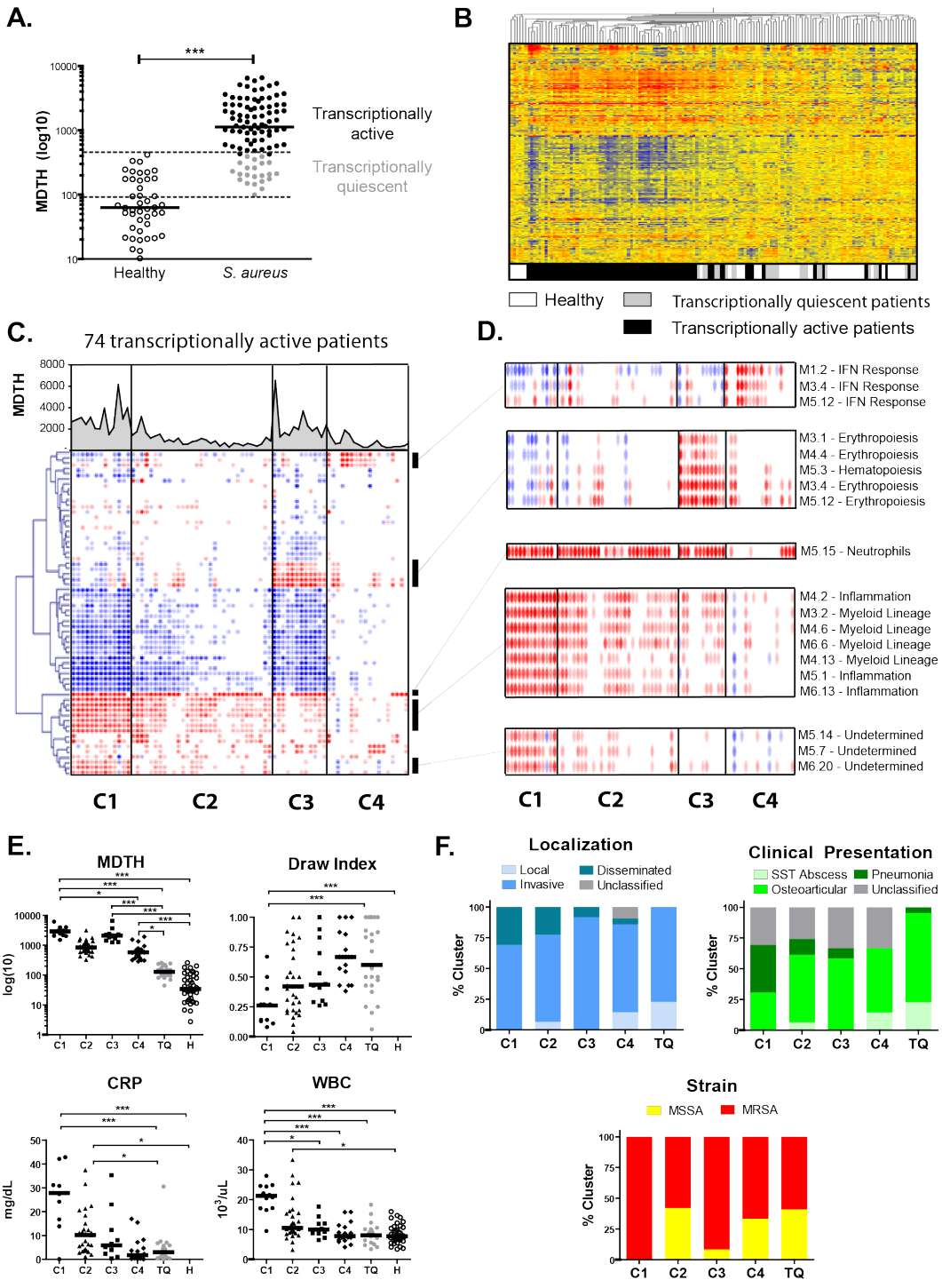

A.
***
MDTH (log10)
Transcriptionally active
Transcriptionally quiescent
Healthy
S. aureus
B
Healthy
Transcriptionally quiescent patients
Transcriptionally active patients
C.
74 transcriptionally active patients
MDTH
C1
C2
C3
C4
D.
M1.2 - IFN Response
M3.4 - IFN Response
M5.12 - IFN Response
M3.1 - Erythropoiesis
M4.4 - Erythropoiesis
M5.3 - Hematopoiesis
M3.4 - Erythropoiesis
M5.12 - Erythropoiesis
M5.15 - Neutrophils
M4.2 - Inflammation
M3.2 - Myeloid Lineage
M4.6 - Myeloid Lineage
M6.6 - Myeloid Lineage
M4.13 - Myeloid Lineage
M5.1 - Inflammation
M6.13 - Inflammation
M5.14 - Undetermined
M5.7 - Undetermined
M6.20 - Undetermined
E.
MDTH
Draw Index
CRP
WBC
log(10)
mg/dL
C1 C2 C3 C4 TQ H
F.
Localization
Local
Invasive
Disseminated
Unclassified
Clinical Presentation
SST Abscess
Osteoarticular
Pneumonia
Unclassified
% Cluster
C1 C2 C3 C4 TQ
Strain
MSSA
MRSA

Figure 17: Individual analysis identifies heterogeneous components of the blood signature to *S. aureus*.

A. Column scatter plot representing the distribution of individual molecular distance to health (MDTH) in healthy controls and *S. aureus* patients. The list of all transcripts composing the modules was used as reference to calculate individual MDTH (***: p<0·001, Mann-Whitney). Horizontal bars represent the group median. Patients with MDTH within healthy range (n=25) were categorized as transcriptionally quiescent (TQ) and represented in grey. **B.** Unsupervised hierarchical clustering of the 10,972 transcripts expressed in at least one of the 143 samples (2-fold normalized, 100 difference in raw data) from the combined training and test sets. **C.** The modular signature was derived for individual transcriptionally active patients (n=74) as compared to the median of the healthy control group for the corresponding patient set (training or test). Four major clusters (C1 through C4) of patients were obtained by K-means clustering and reorganized into a single heatmap, with modules in rows and patients in columns. Molecular distance to health for individual samples is represented as a line chart on top of the heatmap. **D.** Zoom on modules with specific over-expression patterns across the four clusters. **E.** MDTH and clinical lab measurements distribution by cluster. Five or six-group non-parametric ANOVA (Kruskal-Wallis) with Dunn's post-hoc test was applied. (*: p<0·05, **: p<0·01, ***: p<0·001). **F.** Bar charts representing the percent distribution of infection localization, clinical presentation and bacterial strain for the five clusters of patients identified.

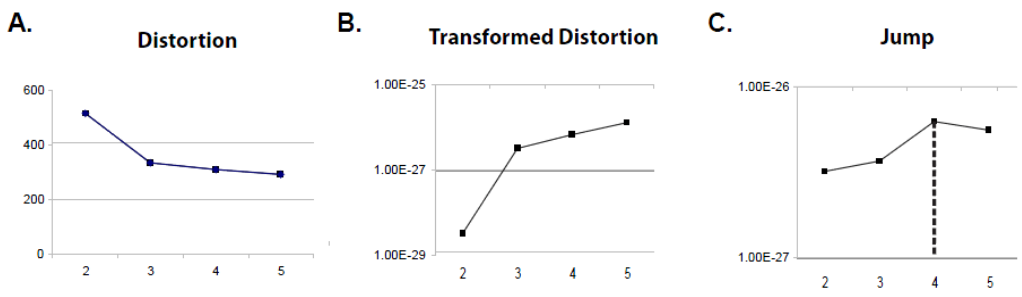


Figure 18: Determination of the best number of K-means clusters from individual patient module expression.

An appropriate K was chosen for clustering this dataset using an information theoretic approach called the "jump method". First, the data was clustered for all K 1 to 5, inclusive. **A.** The distortion of each clustering was calculated. In this instance, we chose to approximate the covariance matrix with the identity matrix so the distortion is simply the mean squared error. **B.** Next, the transformed distortion for each K is calculated by raising the distortion to a power of -(# of dimensions / 2). **C.** Finally, the Jump at each K is calculated as follows: Jump_K = TransformedDistortion_K - TransformedDistortion_K-1. The maximum Jump is used to select K, in this case, K = 4.

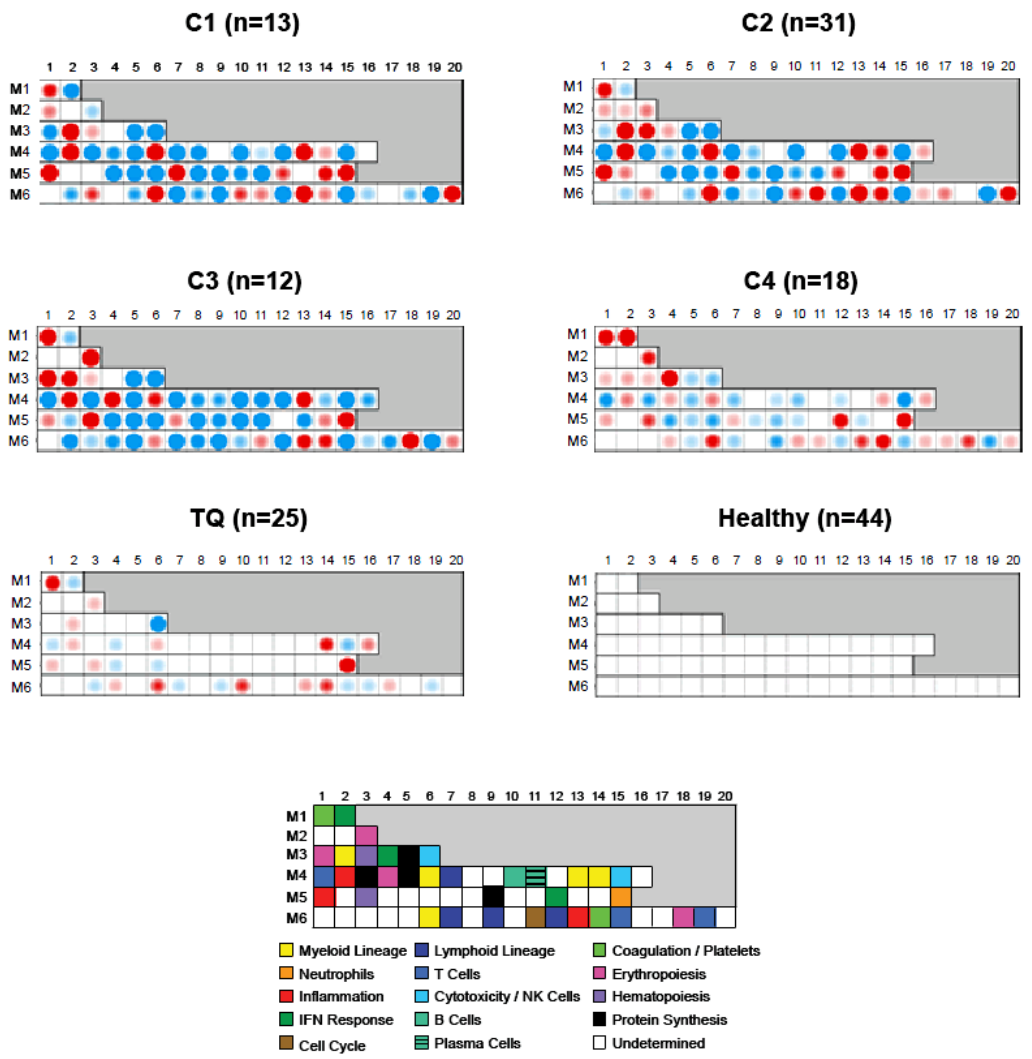


Figure 19: Module signature per cluster of patients.

Module signature was derived for each cluster of molecularly active patients (C1 to C4) and the group of molecularly quiescent (MQ) patients as compared to the combined 44 healthy controls.

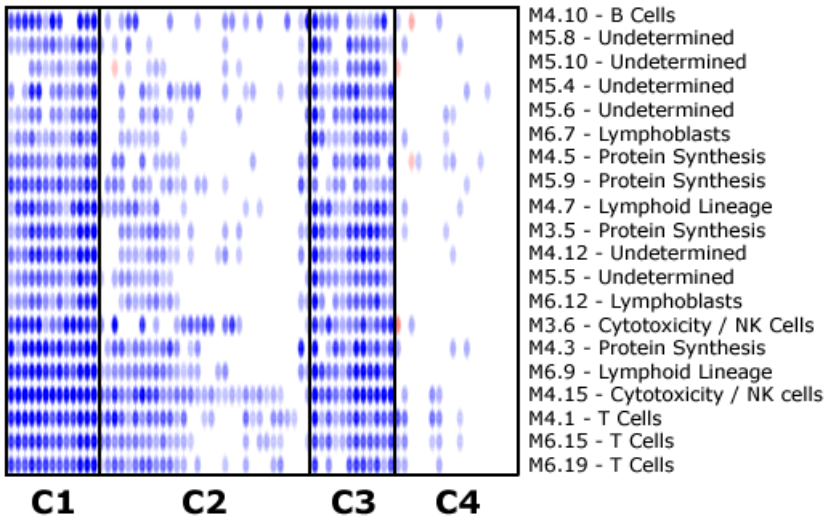


Figure 20: Cluster of under-expressed modules in the 74 molecularly active patients with *S. aureus* infection.

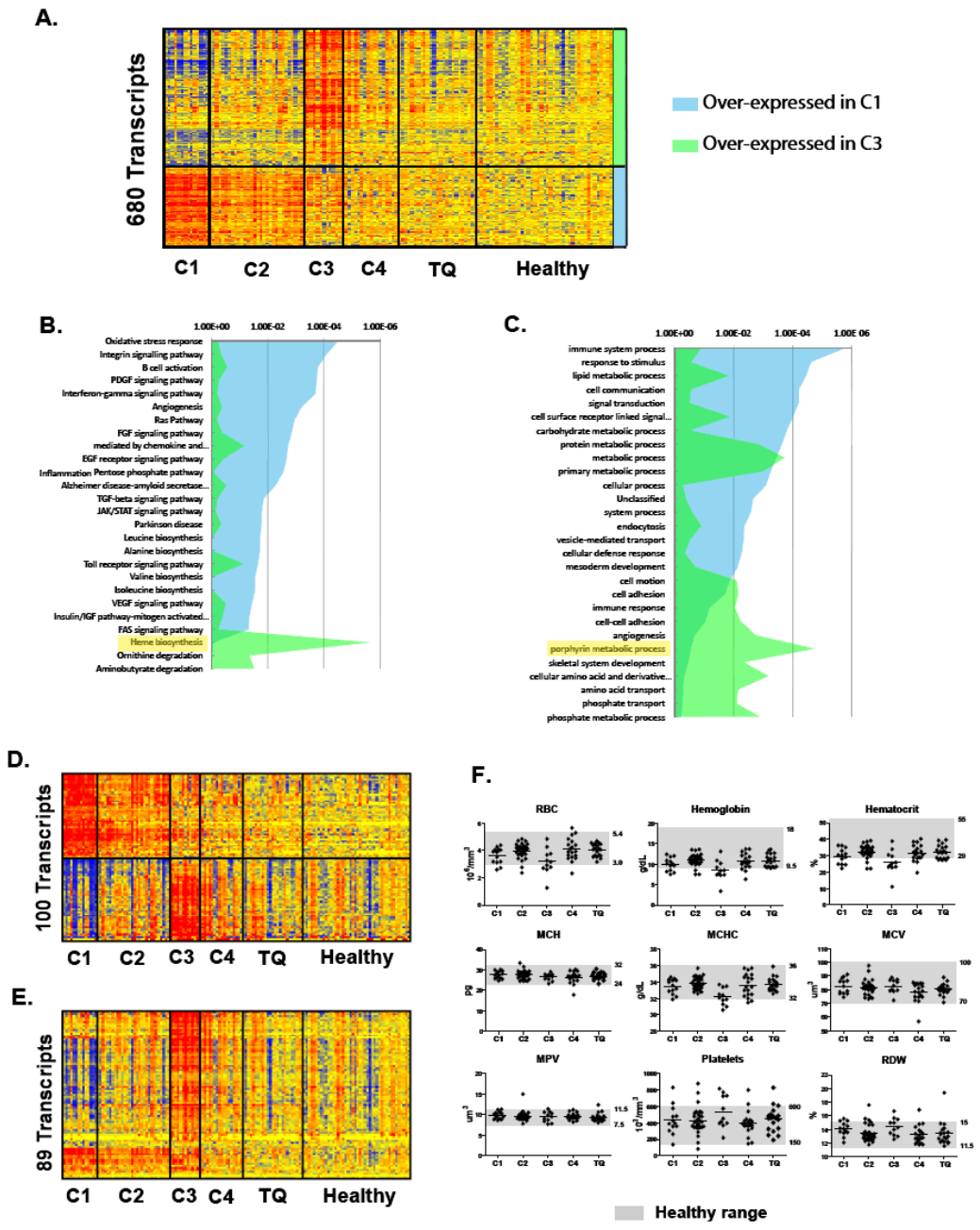


Figure 21: Cluster C3 displays an erythropoiesis signature.

A. Heatmap representing the 680 transcripts differentially expressed between clusters C1 and C3. Statistical comparison between C1 and C3 yielded 680 differentially expressed

transcripts (non-parametric test, α=0·01, Benjamini-Hochberg multiple testing correction, 2x fold change). Of these, 246 transcripts were over-expressed in C1 and 434 transcripts were over-expressed in C3. **B.** Area chart representing pathways significantly enriched ($p<0·05$) in the two gene lists identified in A according to PANTHER (http://www.pantherdb.org). **C.** Area chart representing biological processes significantly enriched ($p<0·01$) in the two gene lists identified in A according to PANTHER. **D.** Heatmap representing the top 100 transcripts over or under-expressed in C3 versus C1. **E.** Heatmap representing the expression of 87 transcripts specifically expressed in CD71+ erythroid precursors in the four patient clusters. **F.** Scatter plot of various CBC measurements for patients organized by transcriptionally active clusters C1 through C4 and transcriptionally quiescent (TQ) patients. The green rectangle represents the range of values for healthy children (two months to 18 years) recorded at Childrens Hospital, Dallas.

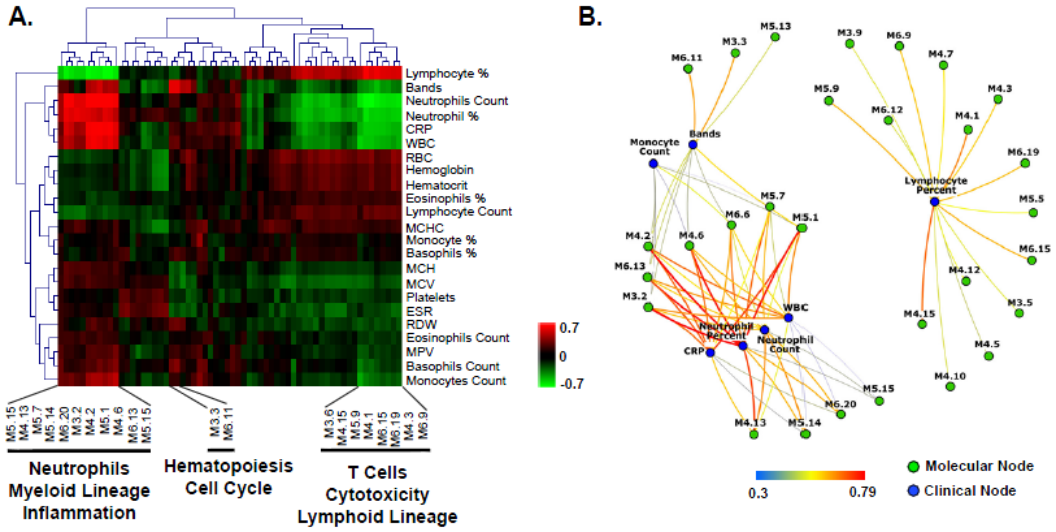


Figure 22: Specific module subsets correlate with laboratory results.

A. Heatmap representing correlation (Spearman R) between module percent expression in columns and continuous laboratory parameters in rows. Hierarchical clustering (Euclidian distance) was applied in both dimensions. **B.** Connection network representing correlation between molecular nodes (modules) in blue and clinical nodes (laboratory parameters) in green. Spearman R correlation greater than 0·3 are represented.

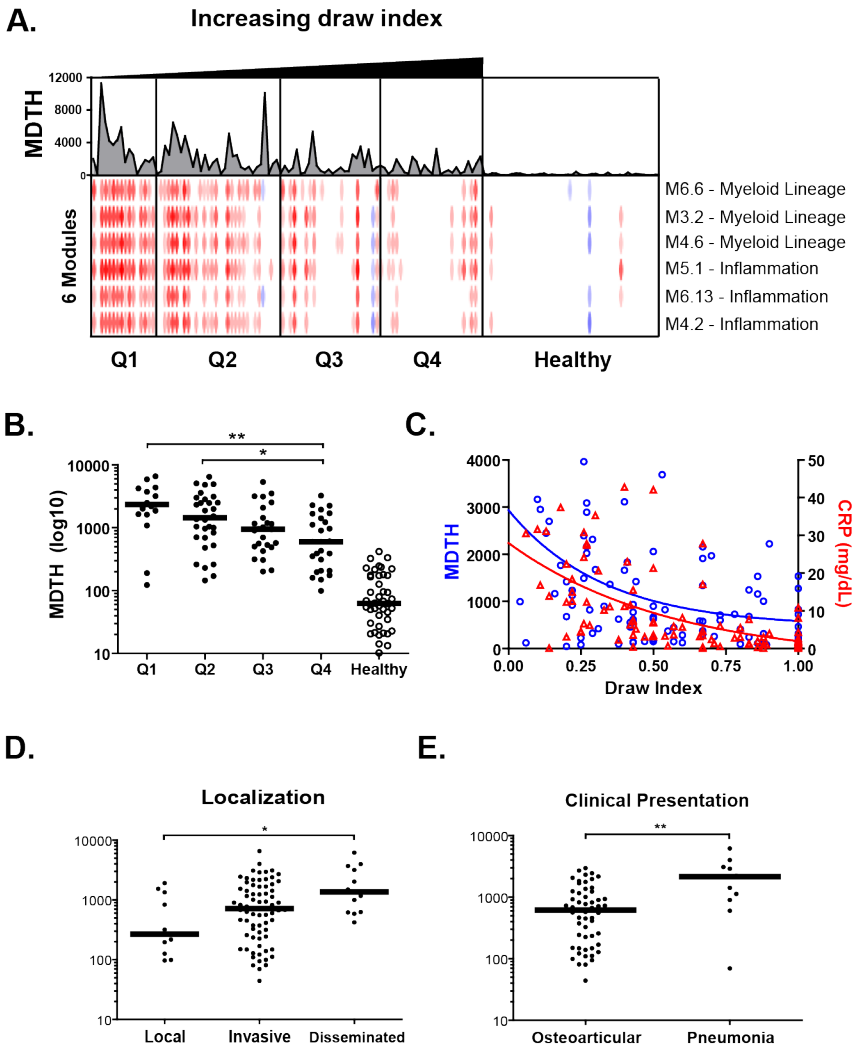


Figure 23: Patient signature varies with blood draw index, dissemination and clinical presentation.

A. Patients were organized in four quarters Q1 through Q4 based on blood draw index (ratio draw day/hospitalization duration). A low draw index signifies proximity to admission while a high draw index signifies proximity to discharge. Transcripts differently expressed between the four draw index quarters were selected by non-parametric ANOVA (Kruskal-Wallis, $p<0{\cdot}01$, Benjamini-Hochberg false discovery rate)

and represented as a heatmap (red, yellow, blue). The same statistical filter was applied at the module level (red, white, blue heatmap below). Individual MDTH was represented above as a line chart. **B.** Column scatter plot of individual MDTH per blood draw index quartile. Horizontal bars represent group median. Non-parametric ANOVA (Kruskal-Wallis) with Dunn's post-hoc test was applied. **C.** Non-linear regression model (one-phase decay) of MDTH (left Y-axis) and CRP (right Y-axis) as a function of blood draw index. **D.** Column scatter plot of individual MDTH per infection localization group. **E.** Column scatter plot of individual MDTH per clinical presentation group. Horizontal bars represent the median value for each group. Non-parametric ANOVA (Kruskal-Wallis) with Dunn's post-hoc test was applied between patient groups.

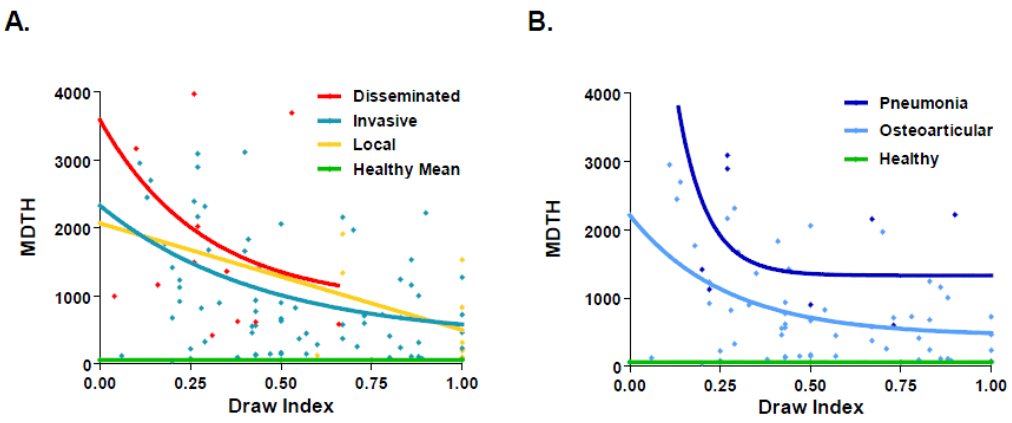


Figure 24: *S. aureus* patient molecular distance to health varies with infection localization, clinical presentation and blood draw index.

A. MDTH as a function of blood draw index for patients grouped by disease localization. **B.** MDTH as a function of blood draw index for patients grouped by disease presentation. A non-linear regression statistical model (one-phase decay) was applied.

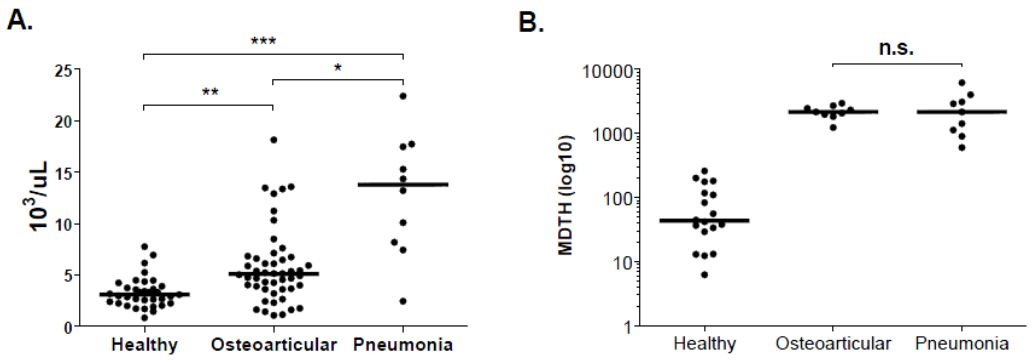


Figure 25: Selection of osteoarticular infection and pneumonia patients for comparison.

A. Scatter plot of neutrophil counts for patients grouped by clinical presentation. Horizontal bars represent the median value for each group. Non-parametric ANOVA (Kruskal-Wallis) with Dunn's post-hoc test was applied. **B.** Column scatter plot of individual MDTH for patients selected for comparison. Horizontal bars represent the median value for each group. Non-parametric test (Mann-Whitney) was applied between the two groups of patients.

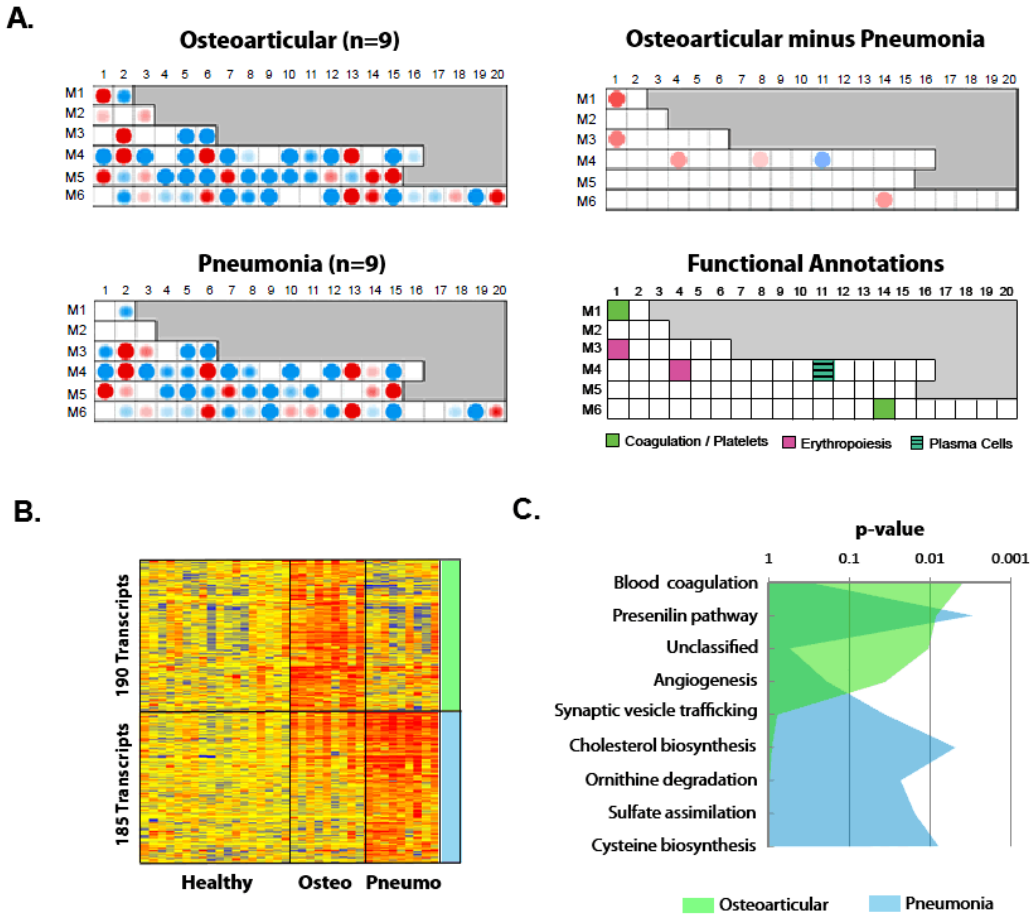


Figure 26: The osteoarticular infection signature displays increased blood coagulation.

We compared the transcriptional signatures from patients with pneumonia and patients with osteoarticular infections. To properly balance osteoarticular and pneumonia groups, patients with pneumonia with a draw index less than 0·75 (nine patients) were selected (active disease). nine patients with osteoarticular infection were selected with matching MDTH so that global quantitative signature was equivalent between the two groups. Nine healthy controls were selected from the training and nine from the test set (18 healthy

controls in total) as reference. **A.** Top left panel: mean module map for the nine patients with osteoarticular infections compared to the 18 healthy controls. Bottom left panel: mean module map for the 9 patients with staphylococcal pneumonia compared to the 18 healthy controls. Top right panel: substraction map of osteoarticular infections minus pneumonia. Only differences greater than 40% are represented. Bottom right panel: Annotation legend for modules identified. **B.** Heatmap representing genes differentially expressed (T-Test, 0·05, no correction) between osteoarticular infections and pneumonia (hierarchical clustering, Pearson). 190 genes were upregulated 1·5-fold or more in osteoarticular infections versus pneumonia and healthy controls. 190 genes were upregulated 1·5-fold or more in pneumonia versus osteoarticular infections and healthy controls. **C.** Area chart representing PANTHER comparison for pathway enrichment between the two lists from C.

Subject #	Age (Yrs)	Race	Sex	Disease	Clinical Presentation															Localization	Strain
					Bacteremia	Osteomyelitis	S. Arthitis	Pyomyositis	SST Abscess	Pneumonia	Empyema	Endocarditis	Cellulitis	Meningitis	Emboli	Sepsis	DVT	PICU	Intubation		
C1_H1181	5	H	M	Healthy																Healthy	n/a
C1_H1183	6	H	F	Healthy																Healthy	n/a
C1_H1239	8	H	M	Healthy																Healthy	n/a
C1_H1242	15	H	F	Healthy																Healthy	n/a
C1_H1245	18	H	M	Healthy																Healthy	n/a
C1_H127	13	AA	F	Healthy																Healthy	n/a
C1_H153	8	H	F	Healthy																Healthy	n/a
C1_H205	5	H	M	Healthy																Healthy	n/a
C1_H213	4	C	F	Healthy																Healthy	n/a
C1_H220	8	O	M	Healthy																Healthy	n/a
C1_H226	7	C	M	Healthy																Healthy	n/a
C1_H229	7	C	M	Healthy																Healthy	n/a
C1_H231	11	AA	F	Healthy																Healthy	n/a
C1_H240	4	H	M	Healthy																Healthy	n/a
C1_H241	7	H	F	Healthy																Healthy	n/a
C1_H275	10	H	F	Healthy																Healthy	n/a
C1_H287	10	AA	M	Healthy																Healthy	n/a
C1_H288	6	AA	F	Healthy																Healthy	n/a
C1_H289	5	AA	M	Healthy																Healthy	n/a
C1_H332	7	C	M	Healthy																Healthy	n/a
C1_H646	4	AA	F	Healthy																Healthy	n/a
C1_H996	5	H	M	Healthy																Healthy	n/a
C1_SA1033	12	C	M	*S. aureus*	■															Unclassified	MRSA
C1_SA1056	3	AA	M	*S. aureus*						■	■									Invasive	MRSA
C1_SA1066	9	AA	M	*S. aureus*		■							■							Invasive	MSSA
C1_SA1069	2	H	F	*S. aureus*	■	■		■	■				■							Invasive	MRSA
C1_SA1070	15	C	M	*S. aureus*	■		■	■	■											Invasive	MSSA
C1_SA1075	12	AA	F	*S. aureus*		■		■	■											Invasive	MRSA
C1_SA1078	8	H	F	*S. aureus*						■										Invasive	MSSA
C1_SA1081	11	AA	M	*S. aureus*	■	■	■	■	■											Invasive	MRSA
C1_SA1083	11	H	M	*S. aureus*	■	■		■	■									■		Invasive	MRSA
C1_SA1109	10	H	F	*S. aureus*	■	■	■	■												Invasive	MRSA
C1_SA1115	0.66	AA	M	*S. aureus*	■	■							■							Invasive	MRSA
C1_SA1117	1.1	AA	M	*S. aureus*	■	■	■		■				■							Invasive	MSSA
C1_SA1118	8	H	F	*S. aureus*			■													Invasive	MRSA
C1_SA1122	5	AA	F	*S. aureus*	■	■	■											■		Invasive	MRSA
C1_SA1141	2	AA	M	*S. aureus*	■	■		■												Invasive	MRSA
C1_SA1142	6	H	F	*S. aureus*	■		■													Invasive	MSSA
C1_SA1152	1.66	AA	F	*S. aureus*	■	■	■						■							Invasive	MSSA
C1_SA1154	10	AA	F	*S. aureus*	■	■			■											Invasive	MSSA
C1_SA1155	11	AA	M	*S. aureus*	■	■	■	■	■											Disseminated	MSSA
C1_SA1204	8	AA	F	*S. aureus*	■	■							■							Invasive	MRSA
C1_SA1228	13	C	M	*S. aureus*	■	■	■		■				■							Invasive	MSSA

C1_SA1257	0.15	AA	F	*S. aureus*	■								■	■		■	■	■	■	Disseminated	MRSA
C1_SA1263	0.66	C	F	*S. aureus*						■	■									Invasive	MRSA
C1_SA1267	1.1	H	F	*S. aureus*						■	■									Invasive	MRSA
C1_SA1276	7	AA	F	*S. aureus*	■	■		■									■			Invasive	MRSA
C1_SA617	11	C	F	*S. aureus*		■														Invasive	MSSA
C1_SA619	17	H	M	*S. aureus*	■	■														Invasive	MSSA
C1_SA632	4	H	F	*S. aureus*	■	■		■		■										Disseminated	MRSA
C1_SA633	6	AA	M	*S. aureus*		■		■	■											Invasive	MRSA
C1_SA907	1.33	H	F	*S. aureus*	■															Unclassified	MRSA
C1_SA908	8.5	AA	M	*S. aureus*	■			■	■	■					■					Disseminated	MSSA
C1_SA909	6	H	F	*S. aureus*	■	■	■	■	■	■					■	■				Disseminated	MRSA
C1_SA943	9	C	M	*S. aureus*		■														Invasive	MRSA
C1_SA944	6	AA	F	*S. aureus*	■	■	■	■					■							Invasive	MRSA
C1_SA949	11	C	M	*S. aureus*	■		■													Invasive	MRSA
C1_SA952	15	AA	M	*S. aureus*	■	■				■	■				■			■		Invasive	MRSA
C1_SA953	5	AA	M	*S. aureus*	■	■	■	■	■				■							Invasive	MRSA
C1_SA954	13	C	M	*S. aureus*	■	■			■											Invasive	MRSA
C1_SA960	3	H	M	*S. aureus*						■	■									Invasive	MRSA
C1_SA999	10	C	F	*S. aureus*	■	■		■	■											Invasive	MRSA

Table 7: Whole blood study: training set subject characteristics

Subject #	Age (Yrs)	Race	Sex	Disease	Clinical Presentation															Localization	Strain
					Bacteremia	Osteomyelitis	S. Arthitis	Pyomyositis	SST Abscess	Pneumonia	Empyema	Endocarditis	Cellulitis	Meningitis	Emboli	Sepsis	DVT	PICU	Intubation		
C2_H1053	0.54	C	F	Healthy																Healthy	n/a
C2_H1191	2.2	H	F	Healthy																Healthy	n/a
C2_H1244	2.2	H	M	Healthy																Healthy	n/a
C2_H1260	2.1	H	M	Healthy																Healthy	n/a
C2_H1261	0.17	H	M	Healthy																Healthy	n/a
C2_H1278	0.75	C	F	Healthy																Healthy	n/a
C2_H1284	0.5	H	F	Healthy																Healthy	n/a
C2_H1285	0.5	O	F	Healthy																Healthy	n/a
C2_H1299	1	AA	M	Healthy																Healthy	n/a
C2_H1321	4	H	F	Healthy																Healthy	n/a
C2_H1353	7	H	M	Healthy																Healthy	n/a
C2_H1354	7	H	F	Healthy																Healthy	n/a
C2_H1355	5	H	F	Healthy																Healthy	n/a
C2_H1357	14	H	M	Healthy																Healthy	n/a
C2_H1358	9	H	M	Healthy																Healthy	n/a
C2_H1359	11	H	M	Healthy																Healthy	n/a
C2_H2003.1	1.42	H	F	Healthy																Healthy	n/a
C2_H924	3	O	F	Healthy																Healthy	n/a
C2_H934	7	C	F	Healthy																Healthy	n/a
C2_H936	3	C	M	Healthy																Healthy	n/a
C2_H938	9	C	M	Healthy																Healthy	n/a
C2_H939	8	C	M	Healthy																Healthy	n/a
C2_SA1006	1.83	AA	M	*S. aureus*					■				■							Invasive	MRSA
C2_SA1007	0.66	AA	M	*S. aureus*					■											Invasive	MRSA
C2_SA1009	0.58	H	M	*S. aureus*	■			■	■	■	■							■	■	Disseminated	MRSA
C2_SA1024	0.16	AA	M	*S. aureus*					■				■					■	■	Invasive	MRSA
C2_SA1030	11	C	M	*S. aureus*	■	■		■	■				■							Invasive	MRSA
C2_SA1032	11	C	M	*S. aureus*	■	■		■	■				■							Invasive	MRSA
C2_SA1076	9	O	F	*S. aureus*			■													Invasive	MRSA
C2_SA1083A	11	H	M	*S. aureus*	■	■		■	■									■		Invasive	MRSA
C2_SA1137	15	H	F	*S. aureus*					■				■							Local	MRSA
C2_SA1176	3.5	AA	M	*S. aureus*		■		■	■				■							Invasive	MRSA
C2_SA1184	1.16	H	M	*S. aureus*					■	■	■									Invasive	MSSA
C2_SA1189	10	H	F	*S. aureus*		■	■	■												Invasive	MSSA
C2_SA1232	6	O	M	*S. aureus*					■				■							Local	MRSA
C2_SA1287	12	C	M	*S. aureus*	■	■	■	■	■											Disseminated	MRSA
C2_SA1306	1.5	C	F	*S. aureus*	■	■														Invasive	MSSA
C2_SA1309	11	C	M	*S. aureus*	■	■		■					■							Invasive	MRSA
C2_SA1313	8	C	M	*S. aureus*		■		■	■				■							Invasive	MRSA
C2_SA1323	16	C	M	*S. aureus*	■	■		■					■							Invasive	MSSA
C2_SA1324	8	AA	F	*S. aureus*	■	■			■											Invasive	MRSA
C2_SA1335	15	C	F	*S. aureus*	■	■	■	■	■				■		■	■		■	■	Disseminated	MRSA
C2_SA1351	2	H	F	*S. aureus*	■	■														Invasive	MRSA

C2_SA493	14	O	M	*S. aureus*														Invasive	MSSA
C2_SA537	9	O	F	*S. aureus*														Invasive	MSSA
C2_SA566	14.5	H	M	*S. aureus*														Invasive	MRSA
C2_SA573	14.5	H	M	*S. aureus*														Invasive	MRSA
C2_SA598	0.25	AA	F	*S. aureus*														Invasive	MRSA
C2_SA614	12.7	O	F	*S. aureus*														Invasive	MRSA
C2_SA623	2.5	H	F	*S. aureus*														Invasive	MSSA
C2_SA626	5	AA	M	*S. aureus*														Disseminated	MRSA
C2_SA634	6.5	AA	M	*S. aureus*														Invasive	MRSA
C2_SA649	9	C	F	*S. aureus*														Invasive	MSSA
C2_SA657	8	AA	F	*S. aureus*														Invasive	MRSA
C2_SA704	6	H	M	*S. aureus*														Invasive	MSSA
C2_SA756	2	H	F	*S. aureus*														Local	MRSA
C2_SA769	1.33	O	M	*S. aureus*														Local	MSSA
C2_SA796	1.58	AA	M	*S. aureus*														Disseminated	MRSA
C2_SA807	1.58	AA	M	*S. aureus*														Disseminated	MRSA
C2_SA811	3	O	M	*S. aureus*														Invasive	MSSA
C2_SA815	1.83	AA	M	*S. aureus*														Disseminated	MRSA
C2_SA824	0.06	C	M	*S. aureus*														Invasive	MSSA
C2_SA844	11	C	F	*S. aureus*														Invasive	MSSA
C2_SA847	1.5	AA	F	*S. aureus*														Invasive	MRSA
C2_SA849	1.08	H	F	*S. aureus*														Invasive	MRSA
C2_SA850	7	H	F	*S. aureus*														Invasive	MSSA
C2_SA855	7.5	AA	F	*S. aureus*														Invasive	MRSA
C2_SA857	4.5	H	F	*S. aureus*														Local	MRSA
C2_SA858	13	C	M	*S. aureus*														Invasive	MSSA
C2_SA859	14	H	M	*S. aureus*														Local	MRSA
C2_SA862	15	AA	F	*S. aureus*														Invasive	MRSA
C2_SA864	5	H	M	*S. aureus*														Invasive	MRSA
C2_SA890	1.5	H	F	*S. aureus*														Invasive	MSSA
C2_SA891	2.5	AA	M	*S. aureus*														Local	MSSA
C2_SA897	10	C	F	*S. aureus*														Local	MRSA
C2_SA906	5	C	M	*S. aureus*														Disseminated	MRSA
C2_SA910	5	AA	M	*S. aureus*														Invasive	MSSA
C2_SA919	14	AA	F	*S. aureus*														Local	MRSA
C2_SA945	7	H	M	*S. aureus*														Local	MRSA
C2_SA985	5	C	M	*S. aureus*														Invasive	MRSA
C2_SA992	10	AA	F	*S. aureus*														Invasive	MRSA

Table 8: Whole blood study: test set subject characteristics

Patient ID	Set	Cluster	Isolate #	Strain (mecA)	SSCmec	agr type	rep PCR	PVL	TSST	SEA	SEB	SEC	SEH	SEG	ETA	ETB
C1_SA1083	Training	C2	1	MRSA	IV	I	300	+								
C2_SA634	Test	C3	3	MRSA	IV	I	300	+								
C2_SA649	Test	C2	4	MSSA	n/a	III	200	+	+	+	+		+	+		
C2_SA906	Test	C1	5	MRSA	IV	I	300	+								
C2_SA493	Test	C4	6	MSSA	n/a	II	300							+		
C2_SA945	Test	TQ	7	MRSA	IV	I	300	+								
C2_SA857	Test	TQ	8	MRSA	IV	I	300	+								
C2_SA864	Test	C1	9	MRSA	IV	I	300	+								
C1_SA619	Training	C2	12	MSSA	n/a	II	300							+	+	
C2_SA614	Test	C1	14	MRSA	IV	I	300	+								
C2_SA815	Test	C3	18	MRSA	IV	I	300	+								
C1_SA633	Training	C2	21	MRSA	IV	I	300	+								
C2_SA897	Test	C2	25	MRSA	IV	I	300	+								
C2_SA859	Test	C4	26	MRSA	IV	I	300	+								
C2_SA849	Test	C1	27	MRSA	IV	I	500	+								
C2_SA626	Test	C1	31	MRSA	IV	I	300	+								
C2_SA657	Test	C2	37	MRSA	IV	I	300	+								
C2_SA598	Test	C4	40	MRSA	IV	I	500	+								
C2_SA573	Test	TQ	41	MRSA	IV	I	500	+								
C1_SA632	Training	C2	45	MRSA	IV	I	300	+								
C2_SA855	Test	C4	51	MRSA	IV	I	300	+								
C1_SA1066	Training	C4	60	MSSA	n/a	I	500	+								
C1_SA953	Training	C4	62	MRSA	IV	I	300	+								
C1_SA617	Training	TQ	63	MSSA	n/a	I	500									
C2_SA847	Test	C4	71	MRSA	IV	I	300	+								
C2_SA756	Test	C4	79	MRSA	IV	I	300	+		+	+		+	+		
C1_SA908	Training	C2	83	MSSA	n/a	I	300	+								
C1_SA909	Training	C4	85	MRSA	IV	I	500	+								
C1_SA952	Training	C1	90	MRSA	IV	I	300	+								
C1_SA1070	Training	TQ	92	MSSA	n/a	II	500									
C1_SA1115	Training	TQ	93	MRSA	IV	I	500	+								
C1_SA907	Training	TQ	94	MRSA	IV	I	300	+								
C1_SA1069	Training	TQ	96	MRSA	IV	I	300	+								
C1_SA1118	Training	C2	98	MRSA	IV	I	300	+								
C1_SA943	Training	C2	101	MRSA	IV	I	300	+								
C1_SA999	Training	C3	103	MRSA	IV	I	500	+								
C1_SA1075	Training	TQ	104	MRSA	IV	I	500	+								
C1_SA1056	Training	C3	108	MRSA	IV	I	300	+								
C1_SA1122	Training	TQ	111	MRSA	IV	I	300	+								
C2_SA992	Test	C3	113	MRSA	IV	I	300	+								
C1_SA960	Training	C1	115	MRSA	IV	I	300	+		+	+		+	+		
C1_SA944	Training	C4	117	MRSA	IV	I	300	+								
C2_SA910	Test	C4	118	MSSA	n/a	I	400							+		
C1_SA1033	Training	C4	120	MRSA	IV	I	700	+						+		
C2_SA1083_A	Test	C1	121	MRSA	-	-	-									
C2_SA1076	Test	C4	122	MRSA	IV	III	1100	+						+		
C2_SA1024	Test	C3	124	MRSA	IV	I	300	+								
C2_SA1007	Test	C1	125	MRSA	IV	I	300	+								

C1_SA949	Training	C2	126	MRSA	IV	I	300	+							
C1_SA954	Training	C2	127	MSSA	n/a	II	400								
C1_SA1228	Training	C2	128	MSSA	n/a	III	200			+					+
C1_SA1154	Training	C3	129	MSSA	n/a	I	300	+							
C2_SA1176	Test	C2	131	MRSA	IV	I	300	+							
C1_SA1155	Training	C2	132	MSSA	n/a	I	300	+							
C1_SA1141	Training	TQ	133	MRSA	IV	III	700	+		+			+	+	
C2_SA1009	Test	C2	134	MRSA	IV	I	500	+							
C2_SA1030	Test	C1	135	MRSA	IV	I	300	+							
C1_SA1081	Training	TQ	137	MRSA	IV	I	300	+			+				
C1_SA1257	Training	C2	138	MRSA	IV	I	300	+							
C2_SA1184	Test	C2	140	MSSA	n/a	I	300	+							
C2_SA1232	Test	TQ	141	MRSA	IV	I	300	+							
C1_SA1117	Training	TQ	142	MSSA	n/a	I	300	+							
C2_SA1006	Test	C4	143	MRSA	IV	I	500	+							

TQ: Transcriptionally Quiescent

Table 9: Characterization of Bacterial Isolates from 63 Patients

Parameter	**TRAINING SET**			**TEST SET**		
	Patients	***Controls***	p^a	***Patients***	***Controls***	p^a
Count	40	22		59	22	
Age (years)	8.00 [0.15-17]	7.0 [4-18]	0.865	6.50 [0.06-16]	3 [0.17-14.00]	0.03
Race	19AA,12H,9C	6AA,11H,4C,1O	0.156	19AA,18H,15C,7O	1AA,13H,6C,2O	0.018
Gender	20M,20F	12M,10F	0.795	34M,25F	11M,11F	0.805
WBC (10^3/mm^3)	9.3 [4.1-28.1]	6.8[3.40-12.6]	0.006	10.10[3.20-33.3]	9.1 [4.20-16.0]	0.065
Neutrophils (%)	52.0 [15-87]	50.7 [34.7-59.8]	0.536	55.00[12.0-84]	35.0 [10.0-55.00]	<0.001
Neutrophils (10^3/mm^3)	5.20 [1.16-22.4]	3.58[1.46-6.94]	0.025	5.63[0.74-18.1]	3.0 [0.84-7.76]	<0.001
Lymphocytes (%)	28.50 [5-67]	37.0 [29.4-53.8]	0.073	33.00 [5.00-72.0]	55.0[33.2-81.0]	<0.001
Lymphocytes (10^3/mm^3)	2.70[0.93-11.0]	2.72 [1.29-4.36]	0.721	3.42 [0.44-13.3]	5.3 [1.44-11.2]	0.016
Monocytes (%)	7.50[1-21]	7.30[5.00-10.6]	0.775	8.0 [2-17]	5.6 [2.00-9.0]	0.096
Monocytes (10^3/mm^3)	0.83 [0.08-3.40]	0.48 [0.19-0.72]	0.036	0.75 [0.14-3.49]	0.44 [0.30-0.98]	0.01
RBC (10^6/mm3)	3.92 [2.6-4.7]	3.90[3.40-4.03]	0.507	4.05 [1.27-5.65]	no data	
Hemoglobin	10.5 [7.3-13.7]	8.80[7.90-11.40]	0.1	10.60 [3.4-13.7]	12.60*	
Hematocrit (%)	31.70 [22.1-38.6]	27.25[25.90-33.30]	0.13	31.70[11.1-40.3]	36.6*	
MCV (um^3)	80.90 [70.7-88.9]	73.50[67.00-86.00]	0.083	81.00 [56.8-97.6]	no data	
MCH (pg)	27.30 [22.9-29.8]	23.75[20.40-29.30]	0.03	27.00 [17.9-33.4]	no data	
MCHC (g/dL)	33.6 [31-35.7]	32.05[30.40-34.20]	0.015	33.5 [30.6-35.9]	no data	
RDW (%)	13.40[11.8-19.4]	12.75[11.50-14.90]	0.337	13.30 [11.6-17.6]	no data	
MPV	9.30 [7.7-15]	7.80[6.50-10.00]	0.009	9.50 [7.8-12.4]	no data	
Platelets (10^3/mm^3)	412 [83-880]	261.00[214.00-288.00]	0.005	414.00 [136-833]	335.00*	
ESR	67.00 [15-142]	no data		68.00 [8 - 135]	no data	
CRP (mg/dL)	5.95 [0.1-37.4]	no data		4.30 [0.1-42.8]	no data	
Hospitalization Duration	11.00 [5-98]	n/a		9.00 [1-53.00]	n/a	
Draw Day	5.00 [2-16]	n/a		5.00 [1-35]	n/a	
Draw Index	0.43 [0.04-1]	n/a		0.50 [0.08-1]	n/a	
Onset Day	4 [1-30]	n/a		3 [1-21]	n/a	
Onset Index	0.64 [0.098-1]	n/a		0.70 [0.14-1]	n/a	

Median values [min-max range]; AA = African-American, C = Caucasian, H = Hispanic, O = Other; M = Male, F = Female; WBC = White Blood Count; RBC = Red Blood Cell count; MCV = Mean Corpuscular Volume; MCH = Mean Corpuscular Hemoglobin; MCHC = Mean Corpuscular Hemoglobin Concentration; RDW = Red blood cell Distribution Width; MPV = Mean Platelet Volume; ESR = Erythrocyte Sedimentation Rate; CRP = C-reactive Protein;

* only 1 sample

Statistics: for categorical variables, Fisher's exact test is used; for continuous variables, Mann-Whitney test is used.

Table 10: Demographic and laboratory characteristics of patients and healthy controls in training and test sets

		Definition	Training (n=40)	Test (n=59)	Total (n=99)
Localization	Local	SST abscess or cellulitis	0 (0%)	10 (16.9%)	10 (10%)
	Invasive	Bacteremia, pneumonia, osteomyelitis, meningitis	33 (82.5%)	41 (69.5%)	74 (74.8%)
	Disseminated	Bacteremia + 2 sites of infection	5 (12.5%)	8 (13.6%)	13 (13%)
	Unclassified		2 (5%)	0 (0%)	2 (2%)
Clinical Presentation	SST Abscess	Skin abscess, no bacteremia	0 (0%)	10 (16.9%)	10 (10%)
	Osteoarticular	Bacteremia and osteomyelitis or suppurative arthritis	29 (72.5%)	27 (45.8%)	56 (56.6%)
	Pneumonia		6 (15%)	5 (8.5%)	11 (11.1%)
	Unclassified		5 (12.5%)	17 (28.8%)	23 (23.2%)

Table 11: Infection localization and clinical presentation distribution for training and test sets

	Up in C1 vs. C3			Up in C3 vs. C1		
#	Symbol	Fold Change	Genbank	Symbol	Fold Change	Genbank
1	NAIP	5.7	NM_004536.2	KRT1	29.3	NM_006121.3
2	SERPINA1	5.4	NM_000295.3	IFIT1L	25.2	NM_001010987.1
3	CR1	4.6	NM_000651.4	TMOD1	24.1	NM_003275.2
4	SLC2A11	4.5	NM_001024938.1	ALAS2	22.6	NM_001037967.1
5	SERPINA1	4.0	NM_001002236.1	HEMGN	20.9	NM_197978.1
6	KREMEN1	4.0	NM_001039571.1	HBD	18.7	NM_000519.3
7	METTL7B	3.9	NM_152637.1	FECH	18.6	NM_000140.2
8	LOC400793	3.9	XM_930558.1	SLC4A1	18.2	NM_000342.2
9	CR1	3.8	NM_000651.4	CA1	18.1	NM_001738.1
10	LOC728519	3.8	XM_001127632.1	BPGM	17.7	NM_001724.3
11	LOC648984	3.8	XM_938063.1	HEMGN	17.7	NM_018437.3
12	SOCS3	3.8	NM_003955.3	GYPB	17.4	NM_002100.3
13	OPLAH	3.7	NM_017570.2	GYPE	15.7	NM_002102.3
14	SLC11A1	3.6	NM_000578.2	FECH	15.2	NM_001012515.1
15	TDRD9	3.6	NM_153046.1	SELENBP1	14.7	NM_003944.2
16	TDRD9	3.5	NM_153046.1	LOC389599	14.3	XM_001131588.1
17	BTNL9	3.5	NM_152547.3	IFI27	13.6	NM_005532.3
18	MGAM	3.5	NM_004668.1	TNS1	13.5	NM_022648.3
19	LOC642684	3.5	XM_926137.1	OR2W3	13.4	NM_001001957.2
20	MCTP2	3.4	NM_018349.2	ITLN1	13.0	NM_017625.2
21	ITGAX	3.4	XM_001127869.1	FBXO7	13.0	NM_001033024.1
22	LOC642103	3.4	XM_936233.1	BCL2L1	12.9	NM_138578.1
23	DERL2	3.4	NM_016041.3	EPB42	12.8	NM_000119.1
24	C19orf59	3.4	NM_174918.2	SLC14A1	12.6	NM_015865.2
25	BCAT1	3.3	NM_005504.4	ALS2CR2	12.5	NM_018571.5
26	TNNI2	3.3	NM_003282.2	HBZ	12.5	NM_005332.2
27	MGST1	3.3	NM_020300.3	RUNDC3A	11.8	NM_006695.3
28	GPR97	3.3	NM_170776.3	SNCA	11.5	NM_000345.2
29	LOC153561	3.2	NM_207331.2	TRIM10	10.9	NM_052828.1
30	MS4A4A	3.2	NM_148975.1	40610	10.3	NM_001002265.1
31	FUT7	3.2	NM_004479.2	ALAS2	10.3	NM_001037968.1
32	COL7A1	3.2	NM_000094.2	SLC6A8	10.1	NM_005629.1
33	IL18R1	3.1	NM_003855.2	PRDX2	10.0	NM_181738.1
34	GRB10	3.1	NM_005311.3	40610	10.0	NM_145021.4
35	PNPLA1	3.1	NM_173676.1	KLF1	9.7	NM_006563.2
36	SLC24A4	3.1	NM_153648.2	ERAF	9.7	NM_016633.2
37	COL4A3BP	3.1	NM_005713.1	ARL4A	9.6	NM_001037164.1
38	MAP3K3	3.1	NM_203351.1	HMG2L1	9.5	NM_005487.3
39	SSH1	3.1	NM_018984.2	GMPR	9.5	NM_006877.2
40	IL1R2	3.1	NM_173343.1	VWCE	9.5	NM_152718.2
41	ST14	3.1	NM_021978.2	KEL	9.5	NM_000420.2
42	TLR5	3.0	NM_003268.4	BPGM	9.2	NM_001724.3

43	GPR97	3.0	NM_170776.3	C14orf45	9.2	NM_025057.1
44	APOB48R	3.0	NM_018690.2	SLC6A10P	9.1	NR_003083.2
45	MANSC1	3.0	NM_018050.2	RSAD2	9.0	NM_080657.4
46	FKBP5	3.0	NM_004117.2	TRIM10	8.9	NM_006778.2
47	GPR141	3.0	NM_181791.1	TTC25	8.8	NM_031421.2
48	DNAJC3	3.0	NM_006260.2	DPM2	8.8	NM_003863.2
49	PAQR6	2.9	NM_198406.1	TRIM58	8.8	NM_015431.3
50	CR1	2.9	NM_000573.3	MYL4	8.8	NM_002476.2

Table 12: Top 50 over-expressed genes in C1 vs. C3 (left) and C3 vs. C1 (right)

	C1			C2			C3			C4			TQ			Kruskal-Wallis Test	
	Median	Min	Max	Median	Min	Max	Median	Min	Max	Median	Min	Max	Median	Min	Max	Chi Square	Pr > Chi-Square
WBC (/ µL)	21300	9500	28000	10550	3200	33300	10000	6500	17700	7800	4100	15900	8100	3400	18300	33.23	<0.0001
Neutrophils (%)	79	37	84	59	40	87	62.6	41	73	47	12	71	43	12	78	35.56	<0.0001
Neutrophils (/uL)	14364	6745	22400	6230	1792	17733	5904	3672	12921	3807	735	11218	3198	1155	10336	46.21	<0.0001
Lymphocyte (%)	12	5	21	25.3	6	40	24	17	35	41	16	71	44	13	72	50.01	<0.0001
Lymphocytes (uL)	2064	1065	4641	3240	896	13320	2444.7	1105	4320	3658	2296	10812	3640	442	9333	10.10	0.04
Monocytes (%)	6.2	2	14	9	3	21	6	2	17	7	1	16.2	7	3	16	2.64	0.62
Monocytes (/uL)	1140	330	3486	858	267	3402	720	176	2448	518	81	1896	707	136	1696	14.79	0.005
RBC (10^6/mm3)	3.88	2.6	4.39	4.12	2.37	4.85	3.16	1.27	4.85	3.9	2.33	5.65	4.15	3.17	4.85	12.48	0.014
Hemoglobin (g/dL)	10.1	7.5	12.4	11.1	7.5	13.6	8.4	3.4	13.1	10.4	6.3	13.7	11.1	9	13.7	14.65	0.005
Hematocrit (%)	30.4	22.3	36.3	32.5	22.1	39.2	25.4	11.1	38.9	30.7	19.7	40.3	32.6	27.3	39.7	11.81	0.019
MCV (um^3)	82.5	74.5	91.2	81.25	73.2	97.6	82.9	72.2	88.9	79	56.8	85.3	80.4	70.7	89	4.26	0.37
MCH (pg)	27.5	24.8	30.2	27.85	24.6	33.4	27	23.4	29.1	26.8	17.9	30.2	27	23.3	30.5	3.83	0.43
MCHC (g/dL)	33.6	31.9	34.6	33.75	32.6	35.7	32.1	30.6	33.7	33.4	31.5	35.7	33.6	32.6	35.9	16.40	0.0025
RDW (%)	14.1	12.1	15.6	13.15	11.8	17.6	14.3	12.6	16.7	12.8	11.8	16.9	13	11.6	19.4	12.00	0.017
MPV (um^3)	9.8	8.6	11.3	9.5	7.8	15	9.5	7.7	11.4	9.35	8.2	10.9	9.1	8	12.4	2.95	0.57
Platelets ($1000/mm^3$)	398	136	833	409.5	83	880	453	219	819	379	158	801	456	146	830	3.44	0.49
ESR	80	33	125	70	17.3	124	103.5	15	142	48.5	16	135	65.5	8	135	5.59	0.23
CRP (mg/dL)	27.8	0.1	42.8	10.2	1.1	37.4	5.9	0.4	35.3	1.8	0.1	17	3.05	0.1	30.5	28.97	<0.0001
Hospitalization Duration	22	5	53	12	2	56	13	8	28	7	3	21	8	1	98	16.48	0.0024
Draw Day	3	1	14	5	1	35	7	3	10	5	2	12	5	1	16	6.91	0.14
Draw Index	0.26	0.08	0.67	0.42	0.04	1	0.44	0.26	1	0.67	0.38	1	0.6	0.06	1	23.38	0.0001

* TQ: Transcriptionally Quiescent

Table 13: Clinical parameters for patients grouped by transcriptional clusters

		C1 (n=13)	C2 (n=31)	C3 (n=12)	C4 (n=18)	TQ (n=25)	Total	Fisher's Exact Test	
								Table Probability (P)	Pr <= P
Localization	Local	0 (0)	2 (6.5)	0 (0)	3 (16.7)	5 (20)	10	1.82E-08	**0.04**
	Invasive	9 (69.2)	22 (71)	11 (91.7)	13 (72.2)	19 (76)	74		
	Disseminated	4 (30.8)	7 (22.6)	1 (8.3)	1 (5.6)	0 (0)	13		
	Unclassified	0 (0)	0 (0)	0 (0)	1 (5.6)	1 (4)	2		
Clinical Presentation	SST Abscess	0 (0)	2 (6.5)	0 (0)	3 (16.7)	5 (20)	10	1.79E-11	**0.0079**
	Osteoarticular	4 (30.8)	17 (54.8)	7 (58.3)	9 (50)	18 (72)	55		
	Pneumonia	5 (38.5)	4 (12.9)	1 (8.3)	0 (0)	1 (4)	11		
	Unclassified	4 (30.8)	8 (25.8)	4 (33.3)	6 (33.3)	1 (4)	23		
Strain	MSSA	0 (0)	13 (41.9)	1 (8.3)	5 (27.8)	11 (44)	30	4.60E-06	**0.0073**
	MRSA	13 (100)	18 (58.1)	11 (91.7)	13 (72.2)	14 (56)	69		

*percents are represented in parenthesis

* TQ: Transcriptionally Quiescent

Table 14: Disease localization, clinical presentation and strain distribution by cluster

			Infection Localization						Clinical Presentation						Cluster							
							Fisher's Exact Test						Fisher's Exact Test								Fisher's Exact Test	
			Local	Invasive	Disseminated	Total	Table Probability (P)	Pr <= P	SST Abscess	Osteoarticular	Pneumonia	Total	Table Probability (P)	Pr <= P	C1	C2	C3	C4	TQ	Total	Table Probability (P)	Pr <= P
	# Patients		10	74	13	97	-	-	10	56	11	77	-	-	13	31	12	18	25	99	-	-
	# Isolates		6	46	9	61	-	-	6	37	6	49	-	-	10	18	7	14	14	63	-	-
Antibiotic Resistance	MRSA		6	35	7	48	0.1	0.5	6	26	5	37	0.1	0.4	10	12	6	11	11	50	0.0016	0.3
	MSSA		0	11	2	13			0	11	1	12			0	6	1	3	3	13		
Agr Type	agr I		6	37	9	52	0.1	1	6	29	6	41	0.1	1	9	14	7	12	12	54	2.60E-03	1
	agr II		0	4	0	4			0	3	0	3			0	2	0	1	1	4		
	agr III		0	4	0	4			0	4	0	4			0	2	0	1	1	4		
Genetic Relatedness	USA 200		0	2	0	2	0	0.9	0	2	0	2	0	1	0	2	0	0	0	2	2.20E-07	0.4
	USA 300		6	29	7	42			6	24	5	35			8	14	6	7	8	43		
	USA 400		0	2	0	2			0	1	0	1			0	1	0	1	0	2		
	USA 500		0	10	2	12			0	7	1	8			1	1	1	4	5	12		
	USA 700		0	1	0	1			0	1	0	1			0	0	0	1	1	2		
	USA 1100		0	1	0	1			0	1	0	1			0	0	0	1	0	1		
Toxin Production	PVL	+	6	38	9	53	0.1	0.5	6	30	6	42	0.2	0.5	9	15	7	12	12	55	1.40E-02	0.7
		-	0	7	0	7			0	6	0	6			0	3	0	2	2	7		
	TSST / SEA / SEB / SEC	+	1	5	0	6	0.2	0.5	1	4	1	6	0.2	0.6	1	2	0	1	2	6	0.028	1
		-	5	40	9	54			5	32	5	42			8	16	7	13	12	56		
	SEH / SEG	+	1	8	1	9	0.1	0.5	5	29	5	39	0.2	1	1	3	0	5	1	10	0.0019	0.3
		-	5	37	9	51			1	7	1	9			8	15	7	9	13	52		

* TQ: Transcriptionally Quiescent

Table 15: Distribution of bacterial isolate characteristics by infection localization, clinical presentation and cluster

Parameter	# of XY Pairs	Spearman r	p-value (two-tailed)
Neutrophils (/uL)	**87**	**0.61**	**< 0.0001**
Neutrophil %	**87**	**0.49**	**< 0.0001**
WBC	**93**	**0.48**	**< 0.0001**
CRP (mg/dL)	**87**	**0.43**	**< 0.0001**
Bands	**49**	**0.33**	**0.0205**
RDW (%)	**92**	**0.30**	**0.0035**
Monocytes(/uL)	**87**	**0.24**	**0.0229**
MCV (um3)	92	0.16	0.1277
MPV (um3)	90	0.13	0.2169
ESR	68	0.13	0.2942
Basophils (/uL)	42	0.13	0.4169
Eosinophils (/uL)	72	0.11	0.339
Platelets (1000/mm3)	92	0.02	0.8824
MCH (pg)	92	0.00	0.9741
Monocytes %	87	-0.10	0.3708
Basophils %	42	-0.12	0.4614
Lymphocytes (uL)	87	-0.21	0.0521
Eosinophils %	72	-0.23	0.0542
Hematocrit (%)	**92**	**-0.24**	**0.022**
MCHC (g/dL)	**92**	**-0.26**	**0.0123**
Hemoglobin (g/dL)	**92**	**-0.27**	**0.0092**
RBC (106/mm3)	**92**	**-0.28**	**0.0068**
Onset Index	**96**	**-0.35**	**0.0004**
Draw Index	**99**	**-0.39**	**< 0.0001**
Lymphocyte %	**87**	**-0.60**	**< 0.0001**

WBC: white blood cell count; **CRP**: C-reactive protein; **RDW**: red blood cell distribution width; **MCV**: mean corpuscular volume; **MPV**: mean platelet volume; **ESR**: erythrocyte sedimentation rate; **MCH**: mean corpuscular hemoglobin; **MCHC**: mean corpuscular hemoglobin concentration; **RBC**: red blood cell count;

Table 16: Spearman correlations between MDTH and clinical parameters

	Local			Invasive			Disseminated			Kruskal-Wallis Test	
	Median	Min	Max	Median	Min	Max	Median	Min	Max	**Chi Square**	**Pr > Chi-Square**
WBC (/ µL)	10100	4800	16700	9550	3400	33300	15850	3200	24400	3.52	0.172
Neutrophils (%)	51.35	12	64	54	12	84	59	37	87	2.85	0.241
Neutrophils (/uL)	4913.3	1212	10688	5225.6	735	22400	8177	1792	18056	4.06	0.131
Lymphocyte (%)	37.8	21	72	33	5	71	24	5	46	8.01	**0.018**
Lymphocytes (uL)	3821.7	2040	7272	3080	442	13320	2726	896	4641	3.86	0.145
Monocytes (%)	6.95	3	13.7	8	2	17	9	1	21	0.56	0.757
Monocytes (/uL)	624.8	303	2004	777.6	136	3486	809.6	81	3402	1.70	0.428
RBC (10^6/mm3)	4.4	3.71	5.65	3.9	1.27	5.16	3.86	2.6	4.47	10.02	**0.007**
Hemoglobin (g/dL)	12.25	10.1	13.4	10.4	3.4	13.7	10.85	7.5	12	10.74	**0.005**
Hematocrit (%)	35.05	32.1	40.3	30.7	11.1	39.7	32	22.1	36.3	11.27	**0.004**
MCV (um^3)	80.8	56.8	97.6	80.8	70.7	93.7	83.55	76.2	91.2	3.49	0.175
MCH (pg)	27.95	17.9	33.4	27	22.9	31.6	28.6	25.5	30	6.01	**0.050**
MCHC (g/dL)	34.3	31.5	35.9	33.4	30.6	35.7	33.6	32.5	34.9	3.69	0.158
RDW (%)	12.6	12.2	16.9	13.4	11.6	19.4	13.95	12.6	15.6	6.29	**0.043**
MPV (um^3)	9.8	8.8	12.4	9.3	7.7	11.4	9.8	8.8	15	5.50	0.064
Platelets ($1000/mm^3$)	331.5	191	419	440	83	880	391	136	578	10.02	**0.007**
ESR	37.5	8	76	71	15	142	69.5	17.3	123	10.09	**0.006**
CRP (mg/dL)	2.2	0.1	10.8	5.95	0.1	42.8	20.6	2.3	31.6	8.82	**0.012**
Hospitalization Duration	3	1	9	9.5	5	98	21	13	56	35.88	**<.0001**
Draw Day	3	1	6	5.5	1	16	7	1	35	10.60	**0.005**
Draw Index	1	0.6	1	0.49	0.06	1	0.27	0.04	0.66	23.43	**<.0001**
Onset to Draw	6	4	9	10	4	37	11	3	35	12.21	**0.002**
Onset Index	1	0.75	1	0.66	0.1	1	0.38	0.11	0.61	29.59	**<.0001**

Table 17: Clinical parameters for patients grouped by infection localization

	Osteoarticular			Pneumonia			Mann-Whitney Test	
	Median	Min	Max	Median	Min	Max	**Z**	**Two-sided Pr > |Z|**
WBC (/ µL)	9200	3400	21600	22100	5100	28100	3.83	**0.0001**
Neutrophils (%)	52	15	87	64	37	84	1.28	0.2005
Neutrophils (/uL)	5040	1066	18144	13786	2448	22400	3.43	**0.0006**
Lymphocyte (%)	34	6	71	22.5	5	39	-1.94	0.0522
Lymphocytes (uL)	2810.1	442	9333	3840	1400	10959	0.98	0.3293
Monocytes (%)	7	2	16.2	7	3	13	-0.29	0.7682
Monocytes (/uL)	738	136	1896	1172	357	3315	2.39	**0.0168**
RBC (10^6/mm3)	3.78	1.27	5.16	4.05	2.37	4.39	0.49	0.6257
Hemoglobin (g/dL)	10.4	3.4	13.7	10.5	7.3	12.4	0.39	0.6934
Hematocrit (%)	30.55	11.1	39.7	31.8	22.2	36.3	0.70	0.4816
MCV (um^3)	80.75	70.7	89	81	74.5	93.7	0.50	0.6191
MCH (pg)	27.15	22.9	30.5	27	24.6	31.6	-0.07	0.9476
MCHC (g/dL)	33.5	30.6	35.7	33	31.3	34.6	-1.18	0.2366
RDW (%)	13.3	11.6	19.4	13.7	12	17.6	1.10	0.272
MPV (um^3)	9.3	7.7	11.4	9.5	8.4	10.5	0.16	0.8708
Platelets ($1000/mm^3$)	448	83	830	398	136	880	-0.06	0.9551
ESR	85.5	15	142	54	44	75	-2.08	**0.0378**
CRP (mg/dL)	4.9	0.1	42.1	13.9	0.4	27.8	1.27	0.2043
Hospitalization Duration	10	5	98	11	6	24	0.07	0.945
Draw Day	6	2	16	4	1	12	-0.75	0.451
Draw Index	0.44	0.06	1	0.27	0.08	1	-0.85	0.394
Onset to Draw	10	4	37	9	6	18	-0.47	0.6419
Onset Index	0.64	0.1	1	0.62	0.27	1	-0.47	0.6371

Table 18: Clinical parameters for patients grouped by clinical presentation

CHAPTER 5: DEVELOPMENT OF A MODULAR FRAMEWORK FOR THE ANALYSIS OF ANTIGEN-PRESENTING CELL TRANSCRIPTIONAL RESPONSES TO PATHOGEN *IN VITRO*

Introduction

The analysis of PBMC and whole blood transcriptional programs in patients with *S. aureus* infections helped us identify a global signature of infection characterized by significant enrichment in transcripts linked to innate immunity. This pro-inflammatory signature correlated with increased numbers of circulating myeloid cells, including neutrophils, monocytes and dendritic cells, as detected by correlations with complete blood count and flow cytometry measurements. Patient-by-patient modular fingerprinting also revealed a significant degree of qualitative and quantitative transcriptional heterogeneity between patients, which we could associate with the dissemination of infection, the type of clinical presentation, and the time in the course of the infection.

The modular analytical framework approach helped us interpret the *ex vivo* blood fingerprint from hundreds of patients by reducing data dimensionality from 48,000 transcripts per sample to 28 modules for PBMC and 62 modules for whole blood. Because these modules were derived from heterogeneous cell populations with very distinct transcriptional programs, they lead to a cell-centric interpretation of the data. Thus, in *S. aureus* infection, the signature is characterized by over-expression of the

neutrophil, myeloid and hematopoietic cell modules, and an under-expression of the T and B cell modules. Innate and adaptive immune cells however sense and respond to pathogens through different pathways, based on the receptors involved in sensing threats, as discussed in introduction. The high coverage / low resolution picture provided by PBMC and whole blood modules limits our interpretation of cell-intrinsic mechanisms of response to infection.

To increase the resolution of our analysis on these innate pathways while leveraging the complexity-reducing capacity of the modular approach, I developed a new analytical framework, built in the context of *in vitro*-activated myeloid cells. In this highly controllable experimental set up with limited tissue heterogeneity, monocyte-derived dendritic cells cultured in the presence of various inflammatory cytokines, were subsequently activated for 6 hours with an array of innate immune stimuli targeting TLRs, cytoplasmic receptors and cytokine receptors. Microarrays were obtained and this large dataset was used as reference to generate a new set of innate immunity modules. I characterized these modules using both knowledge-based and data-driven approaches. I then applied this framework to compare the transcriptional responses induced in monocyte-derived dendritic cells by *Staphylococcus aureus* (gram-positive bacterium), *Salmonella enterica* (gram-negative bacterium) and Influenza virus (H1N1).

Results

Development of a modular framework of the analysis of antigen-presenting cells transcriptional networks

In order to study transcriptional networks induced in antigen-presenting cells in response to innate stimuli, we generated a 353 sample microarray dataset (Table 19) from monocyte-derived IL-4 DC, IFN-alpha DC and IL-15 DC activated for 6 hours *in vitro* with stimuli designed to target various PRR and their downstream signaling cascades (Table 20). These three cytokines mimic various inflammatory conditions and result in antigen-presenting cells with distinct functional phenotypes. IL-4 provides a Th2-like environment conducive to the induction of B cell responses and defense against extra-cellular pathogen. IFN-alpha on the other end is used to mimic viral infection conditions, and is thought to favor the induction of Th1 and CD8+ cytotoxic T cell responses. Finally, IL-15 is thought to confer Langerhans-like phenotype to monocyte-derived DC.

We stimulated cells with TLR2, TLR3, TLR4, TLR5, TLR7, TLR8, TLR9, RIG-I, NOD2, IFNA Receptor, TNF Receptor, IL-1 Receptor, IL-10 Receptor, IL-15 Receptor and IL-21 Receptor ligands. We also stimulated cells with whole live viruses, including 4 strains of H1N1 Influenza (H1N1 A/Brisbane/59/2007, H1N1 A/California/04/2009, H1N1 A/Netherlands/602/2009, H1N1 A/PR/8/34), and 2 strains of Human Immunodeficiency Virus (HIV LAV and HIV JRCSF). Additionally, we stimulated cells with two influenza vaccines (Influenza Virus Vaccine Fluzone 2009-2010, and Influenza A H1N1 2009 Monovalent Vaccine, Sanofi). Finally, we stimulated cells with heat-killed bacteria, including 2 gram-negative pathogens (*E. coli* and *P. gingivalis*) and 2 gram-

positive pathogens (*S. aureus* and *L. monocytogenes*). The 6-hour transcriptional profile of these stimulated DC was characterized by microarray. Figure 27 summarizes the development of the DC modular framework.

This reference dataset was actually composed of samples from 6 different experiments run over the course of 3 years (Table 19). While the platform remained the same during this time (Illumina HT12-v3 bead arrays), significant batch effect was observed between datasets, as observed by principal variance component analysis (PVCA). This screening method allows the identification of prominent sources of variability (biological or technical) within microarray datasets. In this case, 46.9% of the total variance was attributed to a sample's dataset of origin (Figure 28A). This was highlighted by elliptical fit of the scatter plot by dataset number (Figure 28B), where 2 datasets clearly segregated away from the 4 other ones. Following the batch effect correction method employed in the *ex vivo* whole blood study (Chapter 4), we applied CombatR to the DC *in vitro* reference dataset to correct for batch effect. This reduced the variance due to dataset number to 0% (Figure 28C), removing the pre-correction segregation on the elliptical fit of the PCA scatter plot (Figure 28D).

After batch correction, the 353-sample reference dataset was filtered for transcripts absent in all samples, as defined by the Illumina signal detection p-value. This yielded 22,916 out of about 48,000 transcripts present in at least one samples, leaving approximately 8.1 million data points to study. A module extraction algorithm adapted from the one described originally for PBMC was applied to this batch-corrected dataset. The algorithm is described in details in the methodology section as well as in Figure 29.

Briefly, the algorithm first normalizes each sample to its medium RPMI control within each of the 6 original datasets considered. It then scores the expression of each transcript for each activation condition, assigning 1 for over-expression, -1 for under-expression and 0 for no change, as compared to the corresponding RPMI-activated control. Clusters of transcripts are then identified within each of the 6 original datasets. The algorithm then builds a matrix that identifies how many times each pair of transcripts co-cluster over these 6 datasets (a score between 0 and 6 is thus assigned to each possible transcript pair). Finally, a network is built, connecting transcripts according to the score obtained in the matrix, and the algorithm goes recursively through the network to identify groups of transcripts most strongly connected. Once a group is identified, it is defined as a module, the transcripts are removed from the list, and the network is rebuilt for further iteration. This approach yielded 180 modules subsequently treated as single transcriptional units, thereby reducing the number of data points to analyze to 63,540.

Functional characterization of modules

Knowledge-based annotations are limited by: i) the bias of the available literature; ii) the bias of the interpreter; iii) the incomplete understanding of the role molecules play in a system. In order to provide an unsupervised data-driven functional interpretation of modules, I classified them according to their level of expression following DC activation with innate immune stimuli.

To this end, hierarchical clustering of modules based on expression across these stimuli was conducted (Figure 30). Two major groups of modules first separated: 44

modules that were globally over-expressed in response to stimulation, and 136 modules that were under-expressed in response to most stimuli. The 44 over-expressed modules (Appendix C) could be further partitioned into 2 sub-groups based on their response to viral stimuli (viruses, nucleid acids) and bacterial stimuli (heat-killed bacteria, bacterial cell wall components) (Figure 31). The modules associated with interferon and antiviral responses clustered together and were over-expressed in response to TLR3 (Poly I:C, Flu viruses), TLR4 (LPS, HKEC), TLR7 (R837, R848, CL097), TLR8 (R848, CL097), RIG:I (Poly I:C LMW-Lyovec) and IFN-alphaR (IFN-alpha) stimuli. The inflammatory and antibacterial response modules formed a second cluster, with over-expression in response to bacterial stimuli and pro-inflammatory cytokines such as IL-1-beta and TNF-alpha.

In addition to identifying clusters of modules with transcriptional and functional connections, this data-driven categorization approach permitted us to assess the signaling quality and quantity of the various stimuli under study. Poly I:C, IFN-alpha and Poly I:C LMW-Lyovec only induced interferon response modules. Conversely, gram-positive stimuli (HKSA, HKLM) and pro-inflammatory cytokines (IL-1beta, TNF-alpha) only induced inflammatory response modules. Gram-negative stimuli induced both interferon and inflammatory response modules. Flu viruses and vaccines induced the interferon response modules and a subset of inflammatory response modules. Finally, stimuli such as CpG, HIV, IL10, IL-15 and IL-21 induced little transcriptional activity in monocyte-derived DC after 6h activation.

The DC development environment influences transcriptional responses to stimuli

To assess the influence of the development environment on monocyte-derived DC, monocytes were cultured in the presence of three distinct cytokines, IL-4, IFN-alpha and IL-15. PCA analysis of all 353 samples revealed strong segregation by cell population, IFN-alpha DC and IL-4 DC showed the most different patterns (Figure 32A). To simplify the analysis and because the IL-15 DC activation dataset was incomplete, I focused on the IL-4 DC versus IFN-alpha DC comparison. Module fingerprinting of IFN-alpha and IL-4 DC revealed increased baseline expression of anti-bacterial (M5.52, M6.56) and inflammatory response modules (M6.27, M.6.56) in IL-4 DC and increased cell maturation in IFN-alpha DC (Figure 32B). Statistical transcript-level analysis yielded 1,136 differentially expressed transcripts between the two populations (Figure 32C). Ingenuity Pathway Analysis (IPA) identified enrichment of pathways related to anti-bacterial response in IL-4 DC (complement system, coagulation system, recognition of bacteria) while IFN-alpha DC were enriched for antigen presentation, DC maturation and NK cells crosstalk pathways (Figure 32D). Transcriptional profiling of PAMP receptor baseline expression identified increased expression of bacterial detectors TLR2, TLR4, TLR5, CD14, NOD2 in IL-4 DC (Figure 32E), highlighting their anti-bacterial potential.

We then compared the modular fingerprint of each DC population after 6-hour challenge with innate stimuli (Figure 33A). Significant differences were observed in response to Poly I:C, IFN-alpha, R837, MDP and flagellin (Figure 33B). IL-4 DC showed reduced signature in response to Poly I:C and R837. Conversely, IFN-alpha DC responded poorly to IFN-alpha, flagellin, IL-1b and MDP stimulation. More subtle differences between cell subsets were also observed, with IL4 DC not displaying over-expression of a subset of IFN-related modules (M6.12, M6.15, M6.21, M6.61, M6.66) in

response to live flu viruses (Figure 33). These observations support the use of this modular framework to characterize the functional capacity of myeloid cell subsets in response to innate stimuli.

Heterogeneous transcriptional responses to whole pathogens

After using well-defined artificial PAMP ligands to characterize the module response to stimulation in a data-driven fashion, I applied this analytical framework to study the *in vitro* DC population response to activation with three clinically relevant pathogens: a virus (Influenza H1N1), a gram-negative bacterium (heat-killed *Salmonella enterica,* HKSE) and a gram-positive bacterium (heat-killed *Staphylococcus aureus,* HKSA). Transcriptional profiles were obtained longitudinally for both IFN-alpha and IL-4 DC, at 1, 2, 6, 12 and 24 hours after stimulation.

To better understand the modular fingerprints induced by these different pathogens, I first determined which modules could be specifically induced by one, two or all three pathogens in the two DC populations at the 6h time point (Figure 34). The pathogen-specific module selection process is detailed in Figure 34A and Figure 34B. Overall, flu mostly induced IFN response modules, while HKSA induced inflammatory and anti-bacterial modules (Figure 34C). The gram-negative HKSE was able to induce both responses, as expected from the TLR4 capability to signal both through MyD88 (inflammation) and TRIF (interferon).

Interestingly, only a few modules were specific for a single pathogen, highlighting the capacity of distinct infectious agents to elicit a broad array of

transcriptional responses. Only M6.37 (CXCL5, STAT3), a module linked to neutrophil chemottraction, was specifically induced by the gram-positive HKSA in IL4 DC. Similarly, only the gram-negative HKSE was able to induce the maturation module M5.54 (CD86) in IFN-alpha DC. The specific IFN response modules M5.53 and M6.61 (UNC93B1, CCL13) induced by flu but not HKSE in IFN-alpha DC also suggest separation of the IFN response through multiple pathways.

As suggested by their baseline transcriptional phenotype, IL-4 DC showed limited response to flu, with only 6 modules induced by the virus, as opposed to 18 modules induced in IFN-alpha DC. Considering the absence of response of IL-4 DC to Poly I:C and R837 (Figure 33B), it is likely that these programs are elicited through TLR3 and/or TLR7.

Heterogeneous kinetics of response to viruses and bacteria

So far, the analysis of DC transcriptional profiles in response to innate stimuli focused on the 6-hour time point. To gain a better understanding of the kinetics of module expression, I obtained the transcriptional profiles from IFN-alpha and IL-4 DC in response to H1N1, HKSE and HKSA at 1h, 2h, 6h, 12h and 24h (Figure 35A). Filtering of the 44 modules over-expressed identified 40 modules that were over-expressed (>=30%) in at least one experimental condition.

Hierarchical clustering of these modules (Figure 35B) clearly separated inflammatory / anti-bacterial modules and interferon / anti-viral modules. Differences in kinetics between pathogen were observed. Bacterial pathogens were able to induce modular fingerprints as early as 1 hour in both IL-4 and IFN-alpha DC (M5.11, M5.49),

while no module was detected before the 6h time point in response to flu activation in either DC population. While we cannot exclude an inoculum or dose effect, this delay in response to viral pathogens is likely due to the need for internalization for detection, while bacterial pathogens can be readily sensed by cell surface receptors.

While IFN-alpha and IL-4 DC displayed very similar transcriptional kinetics in response to HKSE and HKSA, they responded differently to flu. The virus induced a strong IFN response in IFN-alpha DC, which developed between 2h and 6h, and was sustained up until 24h. No effect of the virus on anti-bacterial programs was observed in IFN-alpha DC. In IL-4 DC, flu not only induced a weaker IFN response, but also strongly down-regulated inflammatory and anti-bacterial modules, this being most evident at 12h. This observation supports the ability of viruses to alter anti-bacterial programs through cell-intrinsic mechanisms, which may be relevant in the context of secondary bacterial infections, following flu-like infection.

To better understand the temporal distribution of module activity, I identified major modular kinetic patterns for a single pathogen in IL-4 DC. I chose HKSE, because it could activate both IFN and inflammatory responses, and because these responses were similar in both DC populations. A clustering approach known as the self-organizing tree algorithm (SOTA)[191] identified 11 patterns of transcription over time (Figure 36A). The patterns obtained are represented in Figure 36B and cluster composition is summarized in Table 21. The earliest responses observed, as early as 1 hour after activation, involved modules from the clusters in the first row of Figure 36B. These included inflammatory response modules M5.11, M6.51 and M6.16, containing transcripts such as IL1A, IL1B, TNF, IL6, IL8, IL15 and NFKB, as well as IFN response modules M4.14, M5.49, M5.29,

M5.61 and M6.52, containing transcripts such as IRF1, IFITs, CXCL10, DHX58 and DDX60. The second wave of activation, detectable after 2 hours, includes modules related to maturation (M5.54), T cell activation (M3.4, M4.24), anti-bacterial response (M6.67, M5.52, M6.28) and regulation of inflammation (M5.16, M5.23, M6.43). Finally, clusters #1, #4 and #5 were activated later, starting after 2 hours, and contained modules linked to DC maturation (M5.50, M6.25), T cell activation (M6.8), and regulation of apoptosis (M6.30).

Application of DC modules to ex vivo whole blood samples in human disease

This modular framework was developed from *in vitro* monocyte-derived DC datasets, and applied so far to further decorticate early transcriptional responses of these cells to infectious pathogens *in vitro*. I then asked whether this analytical tool could be used in whole blood *ex vivo* samples to refine our understanding of inflammatory pathways involved in disease pathogenesis. I selected samples from 8 diseases, including viral infections (Influenza A, acute HIV, RSV), bacterial infections (*S. aureus*, melioidosis caused by the gram-negative bacterium *Burkholderia pseudomallei*, acute tuberculosis) and autoimmune diseases (SLE and SoJIA). The sample distribution is summarized in Table 22.

DC module fingerprints were obtained for each disease, using their respective group of healthy controls as reference. The 44 modules that display over-expression in response to *in vitro* stimuli were then filtered to retain 35 modules induced (>=30%) in at least one of the disease conditions. Hierarchical clustering of these modules revealed common and specific innate signals induced in these diseases (Figure 37). Of interest,

most diseases (but *S. aureus*) displayed over-expression of a subset of IFN response modules (M5.29, M4.14, M5.50, M6.52, M4.15, M5.61, M6.26). A second group of IFN modules were only induced in flu, HIV and SLE (M5.57, M5.53, M5.33, M6.12, M6.17), highlighting the further partitioning of the IFN/antiviral response. Some modules were only induced in a single disease, such as M6.35 (IFN-alphas) in flu, M4.20 (inflammation) in RSV, M5.16 (regulation of inflammation) in SoJIA, or M6.30 (regulation of apoptosis) in HIV infection.

Clustering of disease inflammatory profiles revealed degrees of similarity between conditions. Not unexpectedly, the bacteria *B. mallei* and *S. aureus* clustered together, although the gram-negative pathogen could induce a broader IFN response. The acute viral infections cause by influenza and HIV displayed similar profiles, and clustered with SLE. RSV however, did not resemble other viral fingerprints and was the only condition with absent IRF1 (M5.49) or UNC93B (M6.61) induction. Interestingly, the IL-1-mediated rheumatoid disease SoJIA, clustered with tuberculosis, and with monocyte-derived activated DC activated with MDP (data not shown), suggesting the involvement of NOD2 in disease pathogenesis.

Increased frequency of compartmentalization and lipid metabolism transcripts in S. aureus patients

Finally, I selected all modules over-expressed in *S. aureus* infection, and asked whether these patients displayed a specific signature when compared to the seven other diseases considered. 30 modules were selected (≥30%) and hierarchical clustering was

conducted (Figure 38, *ex vivo* panel). The clustering pattern revealed 3 major groups of modules: i) a set of 6 IFN modules, over-expressed in all diseases (lower cluster); ii) a group of 9 modules linked to inflammatory and antibacterial responses, induced in "bacterial" diseases (*S. aureus*, Tuberculosis, Melioidosis) and SoJIA (middle cluster); iii) A group of 15 modules (upper cluster) that are specifically over-expressed in the blood of patients with *S. aureus* infection but not other diseases. These modules were highly enriched in transcripts linked to either cellular compartmentalization (endocytosis, phagocytosis, vesicle trafficking) or lipid metabolism. These *S.aureus*-specific modules were not part of the 44 ones that are consistently over-expressed in response to *in vitro* stimuli, and were not over-expressed in vitro in IL4 DC in response to TLR ligands and HKSA (Figure 38, *in vitro* panel). They seem to represent elements of *S. aureus* pathogenesis that are not reflected in our *in vitro* DC model.

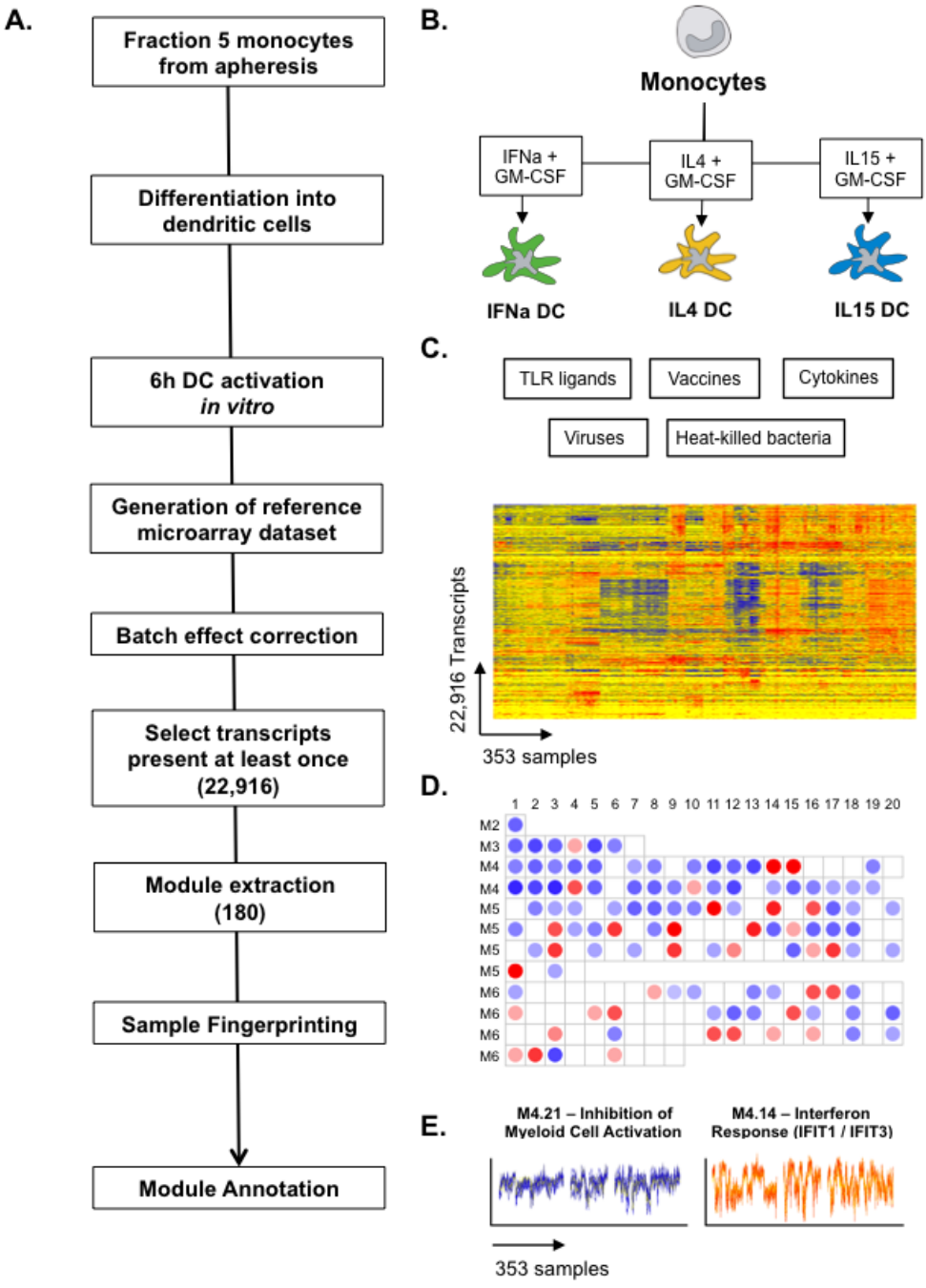


Figure 27: Development of a modular analytical framework to study dendritic cell transcriptional profiles

A. Analytical framework development workflow. Human monocytes obtained from healthy donor apheresis frozen fraction 5 were first differentiated *in vitro* in IL-4 DC, IFN-alpha DC or IL-15 DC. After culture, cells were activated for 6 hours with a variety of innate immune stimuli, including TLR ligands, cytokines, vaccines, viruses and heat-killed bacteria. A reference microarray dataset was generated and a computer algorithm identified 180 transcriptional modules. Module expression for each of the stimuli in each DC population was assessed on an individual basis, and fingerprints were derived. Modules were then annotated using both knowledge-based and data-driven approaches. **B.** Generation of three subsets of monocyte-derived DC. C. Activation of DC by various stimuli and generation of a reference microarray dataset. **D.** Module extraction and fingerprinting of stimuli-induced DC signatures. **E.** Example of modular expression patterns across the 353 reference samples.

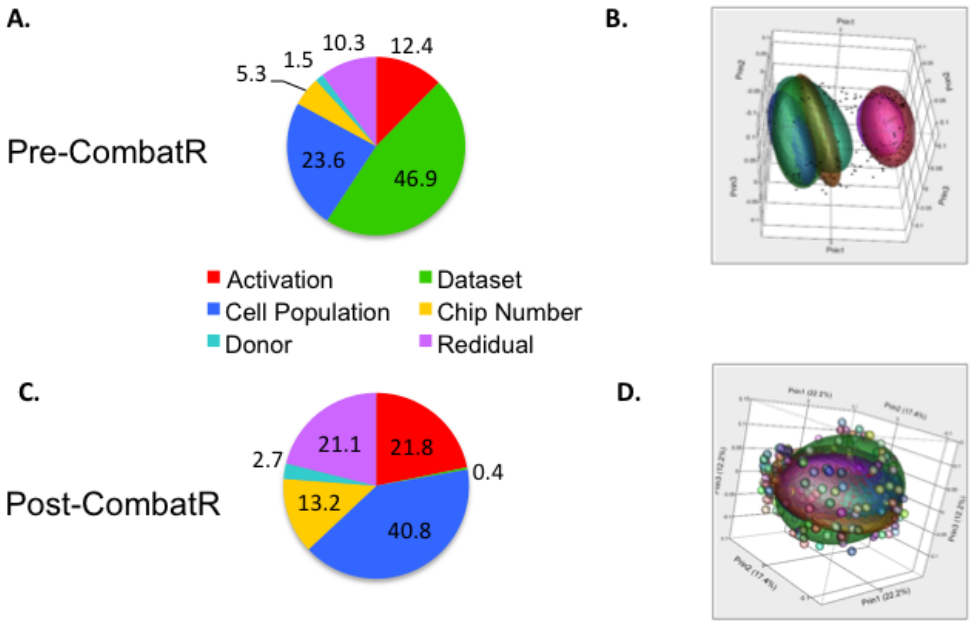
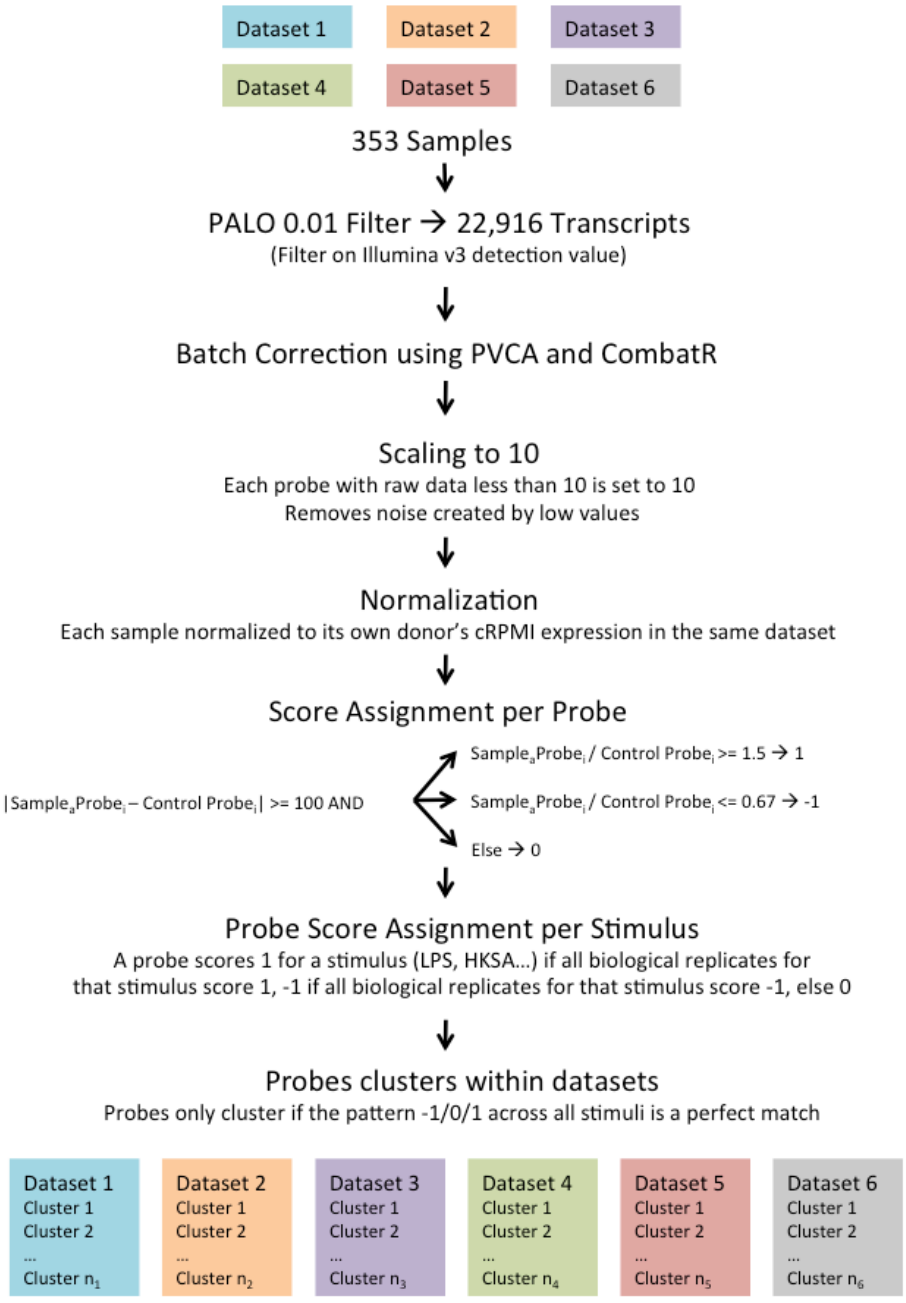


Figure 28: Reference dataset batch correction.

A. Pie-chart representing the results of principal variance component analysis (PVCA) including 5 parameters: activation, dataset, cell population, chip number and donor. Percents are represented. **B.** Elliptical fit of scatter plot representing the segregation of samples by dataset before CombatR correction. **C.** Pie-chart representing the results of PVCA after CombatR correction for the dataset parameter. **D.** Elliptical fit of scatter plot representing the segregation of samples by dataset after CombatR correction.


Dataset 1
Dataset 2
Dataset 3
Dataset 4
Dataset 5
Dataset 6
353 Samples
PALO 0.01 Filter → 22,916 Transcripts
(Filter on Illumina v3 detection value)
Batch Correction using PVCA and CombatR
Scaling to 10
Each probe with raw data less than 10 is set to 10
Removes noise created by low values
Normalization
Each sample normalized to its own donor's cRPMI expression in the same dataset
Score Assignment per Probe
|Sample_aProbe_i – Control Probe_i| >= 100 AND
Sample_aProbe_i / Control Probe_i >= 1.5 → 1
Sample_aProbe_i / Control Probe_i <= 0.67 → -1
Else → 0
Probe Score Assignment per Stimulus
A probe scores 1 for a stimulus (LPS, HKSA…) if all biological replicates for
that stimulus score 1, -1 if all biological replicates for that stimulus score -1, else 0
Probes clusters within datasets
Probes only cluster if the pattern -1/0/1 across all stimuli is a perfect match
Dataset 1
Cluster 1
Cluster 2
…
Cluster n1
Dataset 2
Cluster 1
Cluster 2
…
Cluster n2
Dataset 3
Cluster 1
Cluster 2
…
Cluster n3
Dataset 4
Cluster 1
Cluster 2
…
Cluster n4
Dataset 5
Cluster 1
Cluster 2
…
Cluster n5
Dataset 6
Cluster 1
Cluster 2
…
Cluster n6

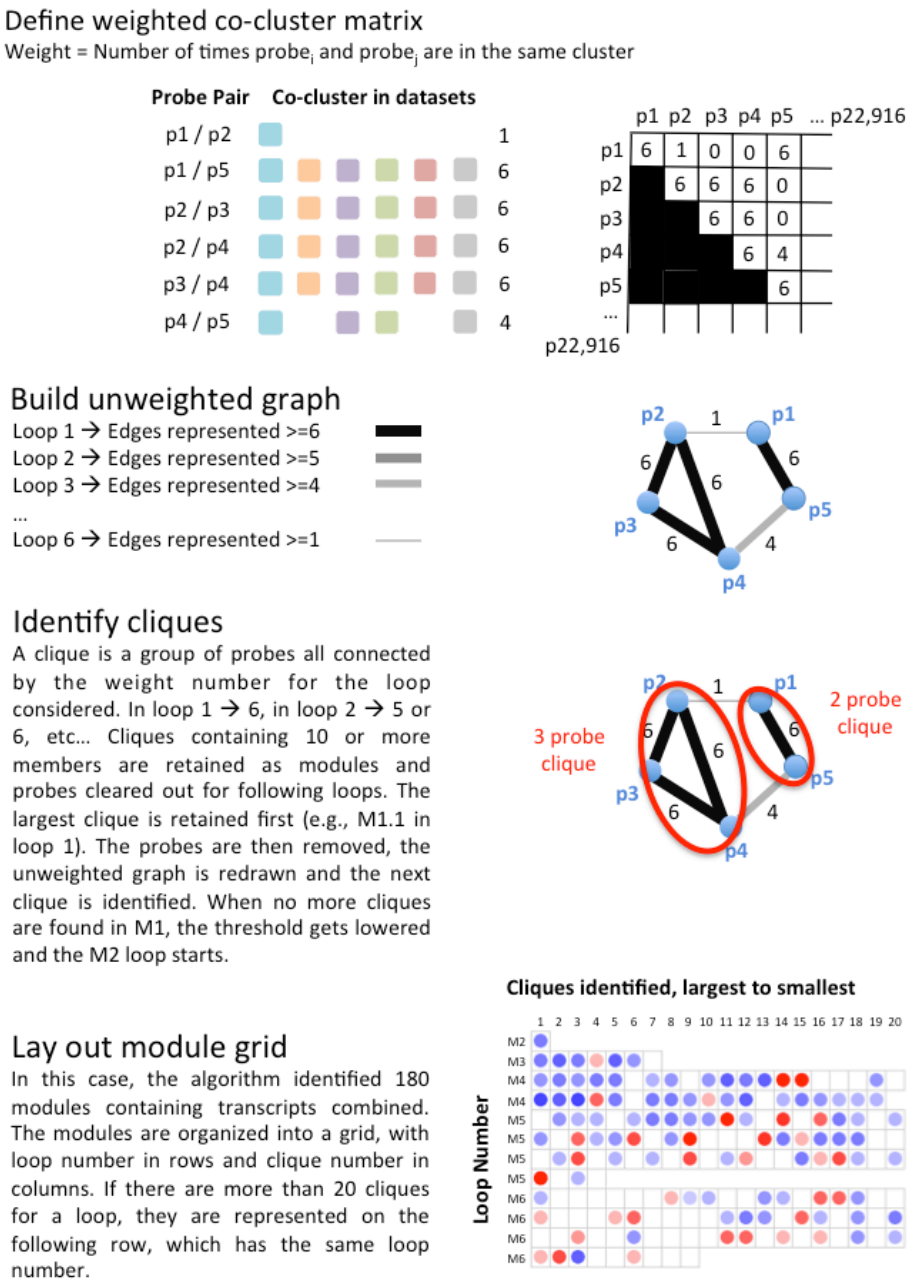


Figure 29: Module construction algorithm.

The details of the algorithm are available in the methods section.

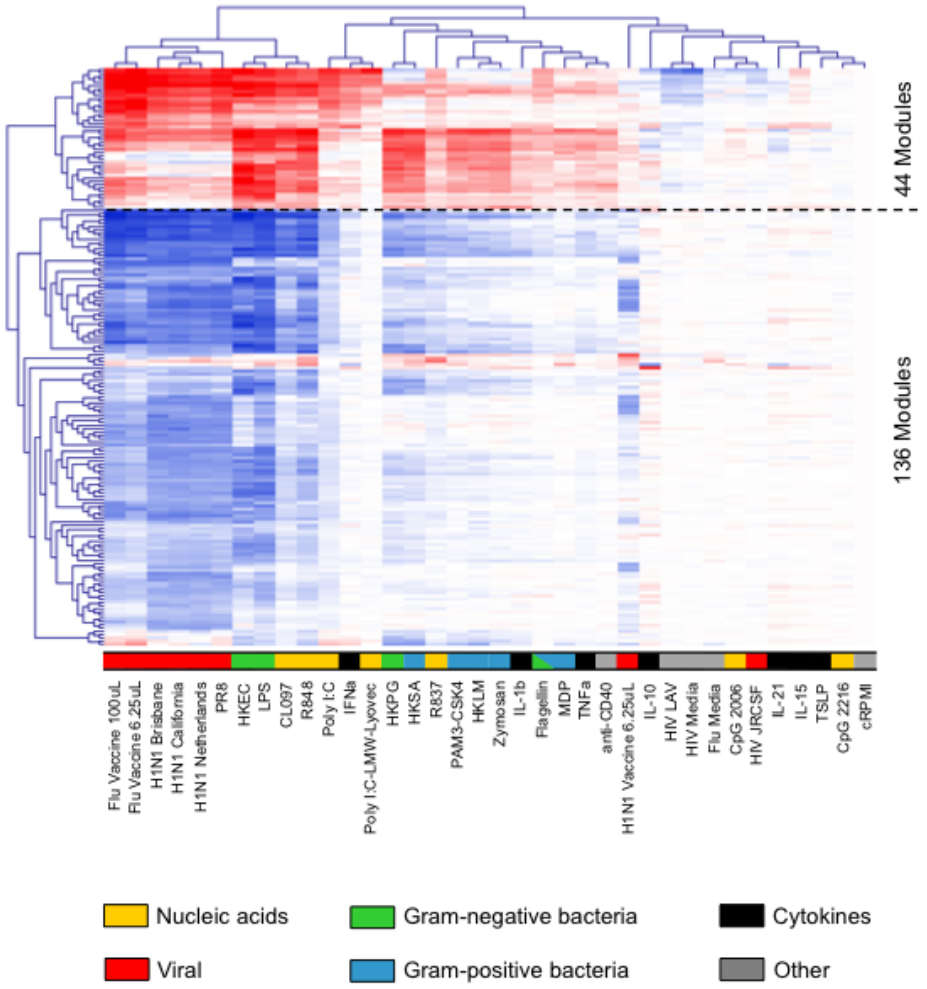


Figure 30: Hierarchical clustering of 180 modules according to their response to innate stimuli.

Module expression was obtained for each of the 36 stimuli considered in this study. Hierarchical clustering (Pearson correlation) was conducted for modules and stimuli. Stimuli were colored according to their category (nucleic acid, virus, vaccine, bacterial cell wall component, gram-positive bacteria, gram-negative bacteria, cytokine, other or control)

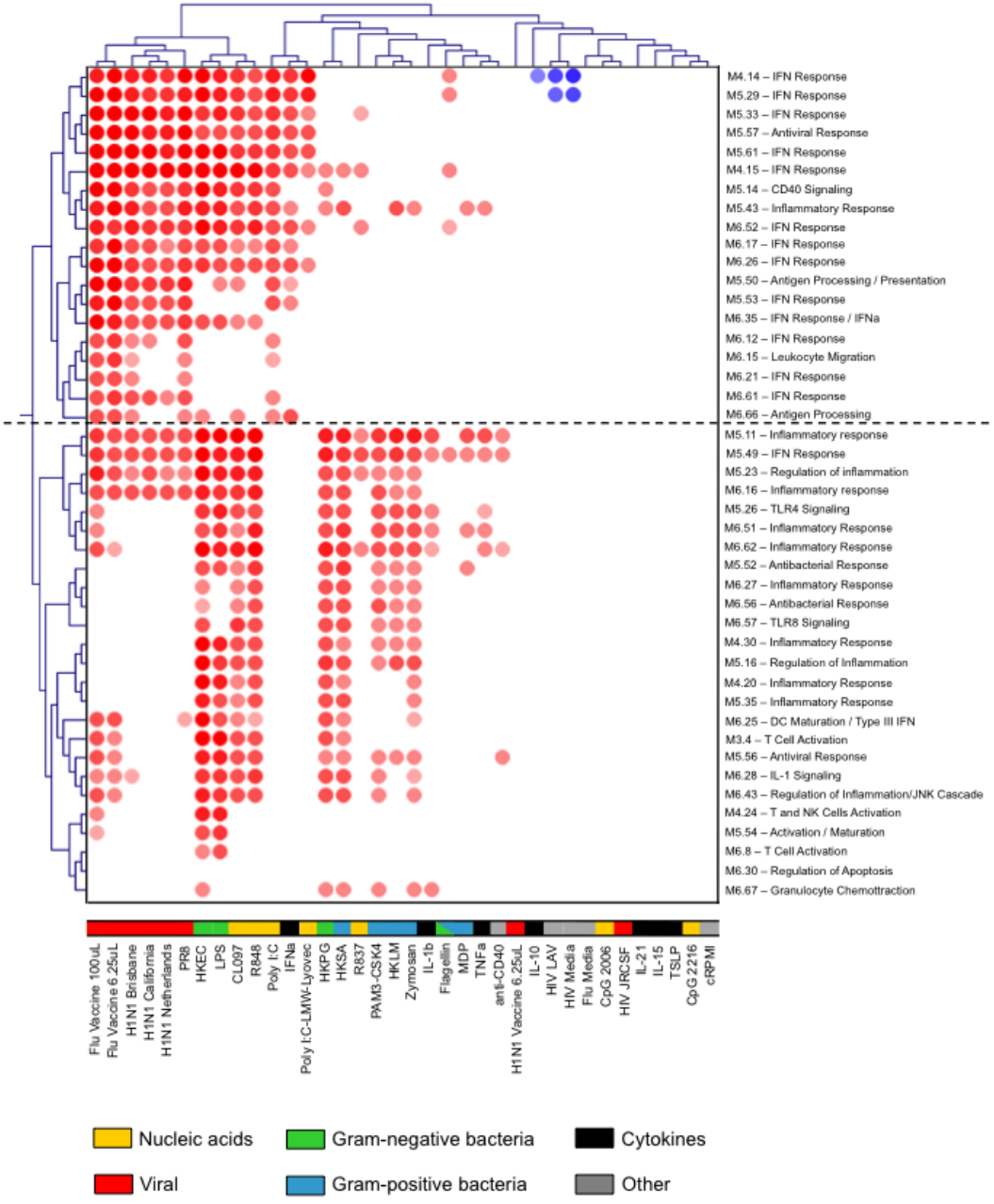


Figure 31: 44 modules over-expressed in response to innate stimuli. DC fingerprint in response to innate stimuli after 6h stimulation.

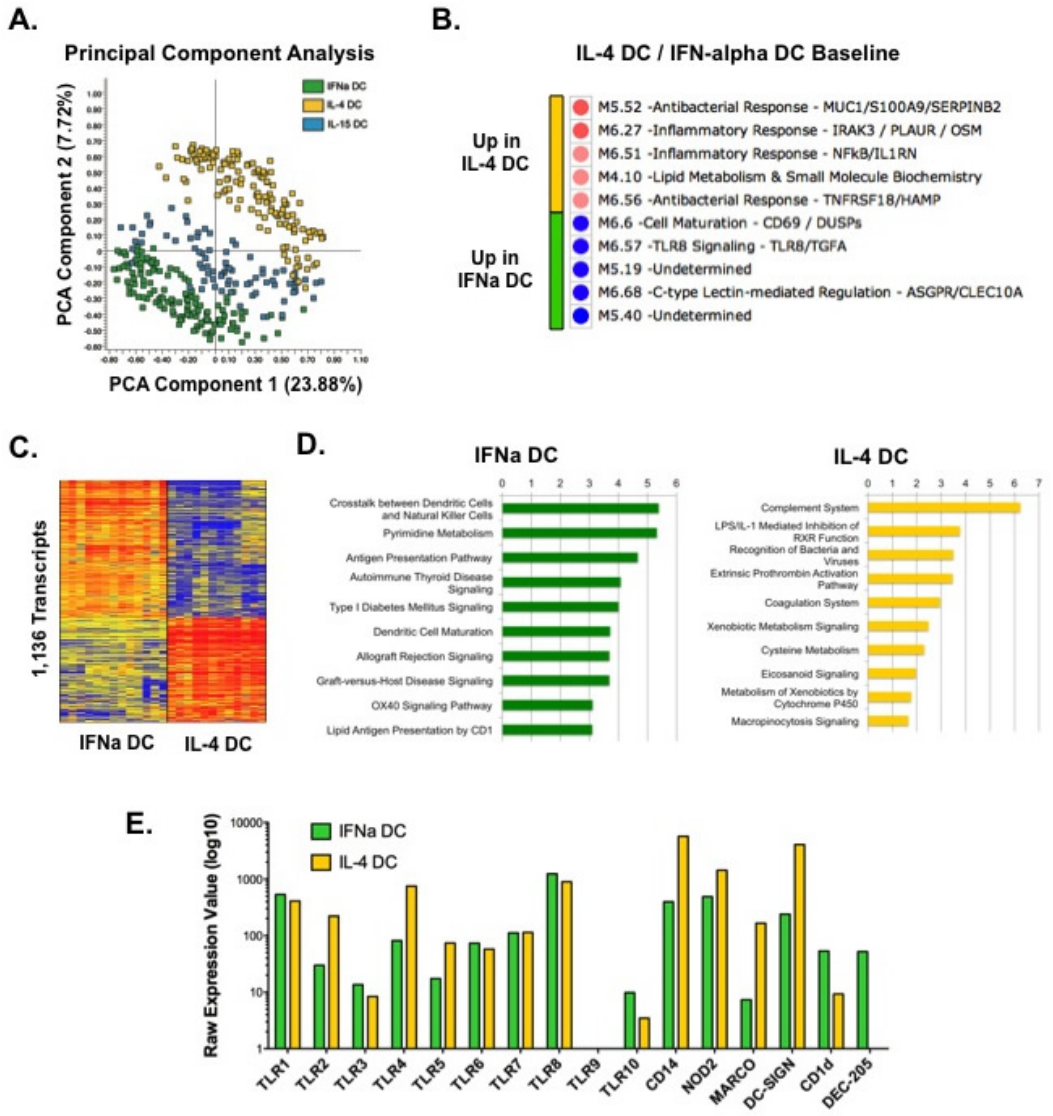


Figure 32: Baseline differences between IFNa DC and IL-4 DC

A. Principal component analysis of the 353 reference samples colored by cell population. **B.** Top 10 modules differentially expressed between IFNa and IL-4 DC. IFNa DC are used as reference. **C.** Heatmap representing the hierarchical clustering of 1,136 transcripts differentially regulated between IFNa and IL-4 DC activated with RPMI

medium for 6h (Welch t-test, p-value cutoff 0.01, Benjamini-Hochberg false discovery rate, 3x up/down). **D.** Top 10 enriched pathways amongst genes over-expressed in IFNa DC (green) and IL-4 DC (yellow) identified by Ingenuity Pathway Analysis. **E.** Bar charts representing the log10 raw expression value of selected PAMPs in IFN-alpha (green) and IL-4 DC (yellow) at baseline.

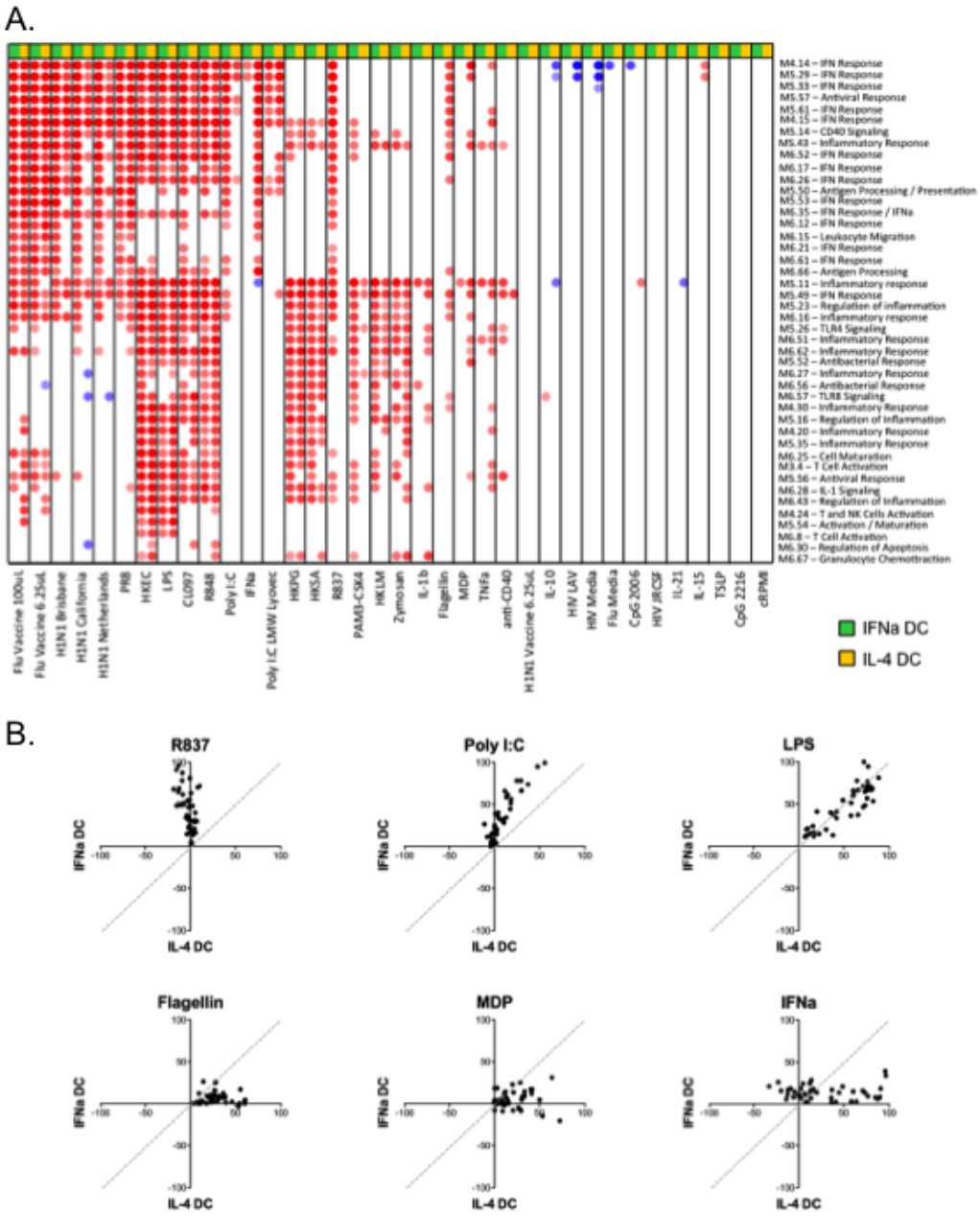


Figure 33: IFNa and IL-4 DC respond differently to viral and bacterial stimuli

A. Modular fingerprints of IFNa and IL-4 DC in response to innate immune stimuli. Modules greater or equal to 30% or less than or equal to -30% are represented. **B.** Scatter plots of module expression data in IL-4 DC (x-axis) and IFNa DC (y-axis) highlighting differences in response to bacterial and viral stimuli between the two cell populations.

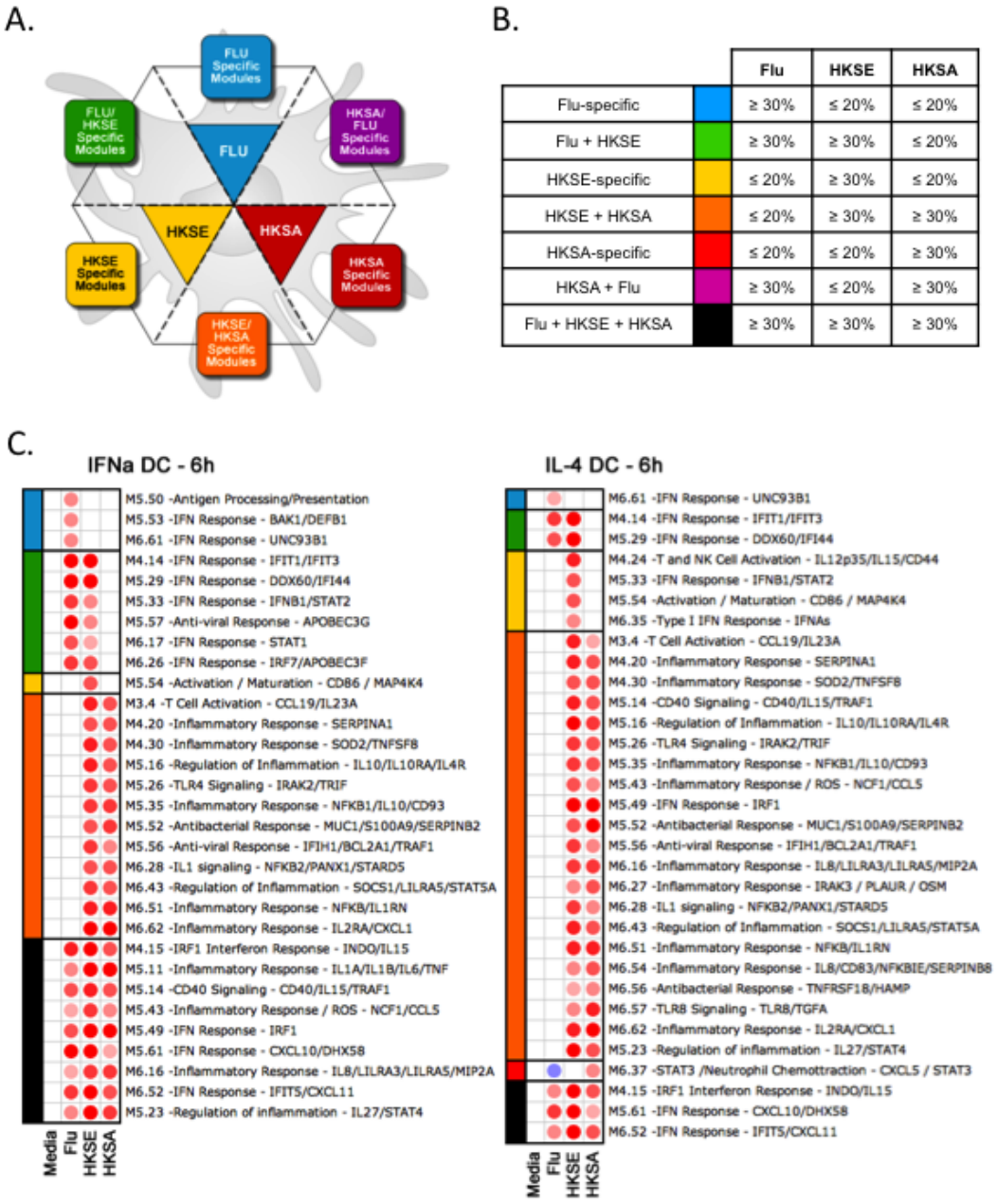


Figure 34: Identification of pathogen-specific 6h modular fingerprints in IFNa DC and IL-4 DC

A. Schema representing the color-schemes associated with module specificity by single or combinations of pathogens. **B.** Table detailing the threshold values selected to classify modules. **C.** Pathogen-specific modular fingerprint for IFNa DC (average of 3 donors). **D.** Pathogen-specific modular fingerprint for IL-4 DC (average of 3 donors).

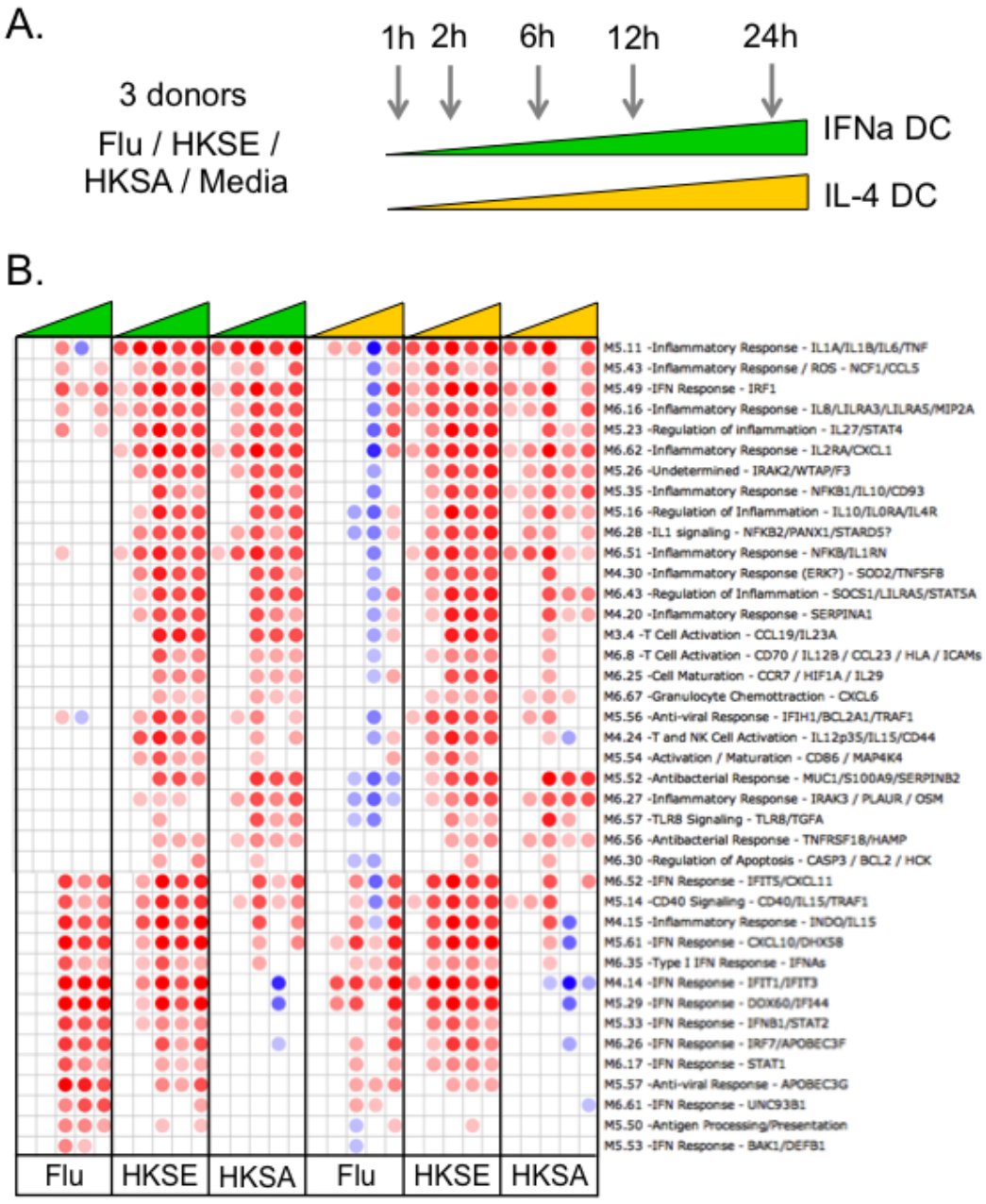


Figure 35: Heterogeneous transcriptional kinetics of IFN-alpha and IL-4 DC in response to pathogen

A. Description of the *in vitro* DC infection time course experiment. IFN-alpha and IL-4 DC from 3 different donors were generated and activated *in vitro* for 1h, 2h, 6h, 12h or 24h with flu (H1N1), HKSE or HKSA. **B.** Modular fingerprints for each DC population, organized by pathogen and ordered by increasing time point. Data represents the average of 3 donors. Each sample is normalized to its own media control (same donor, same cell population, same time point).

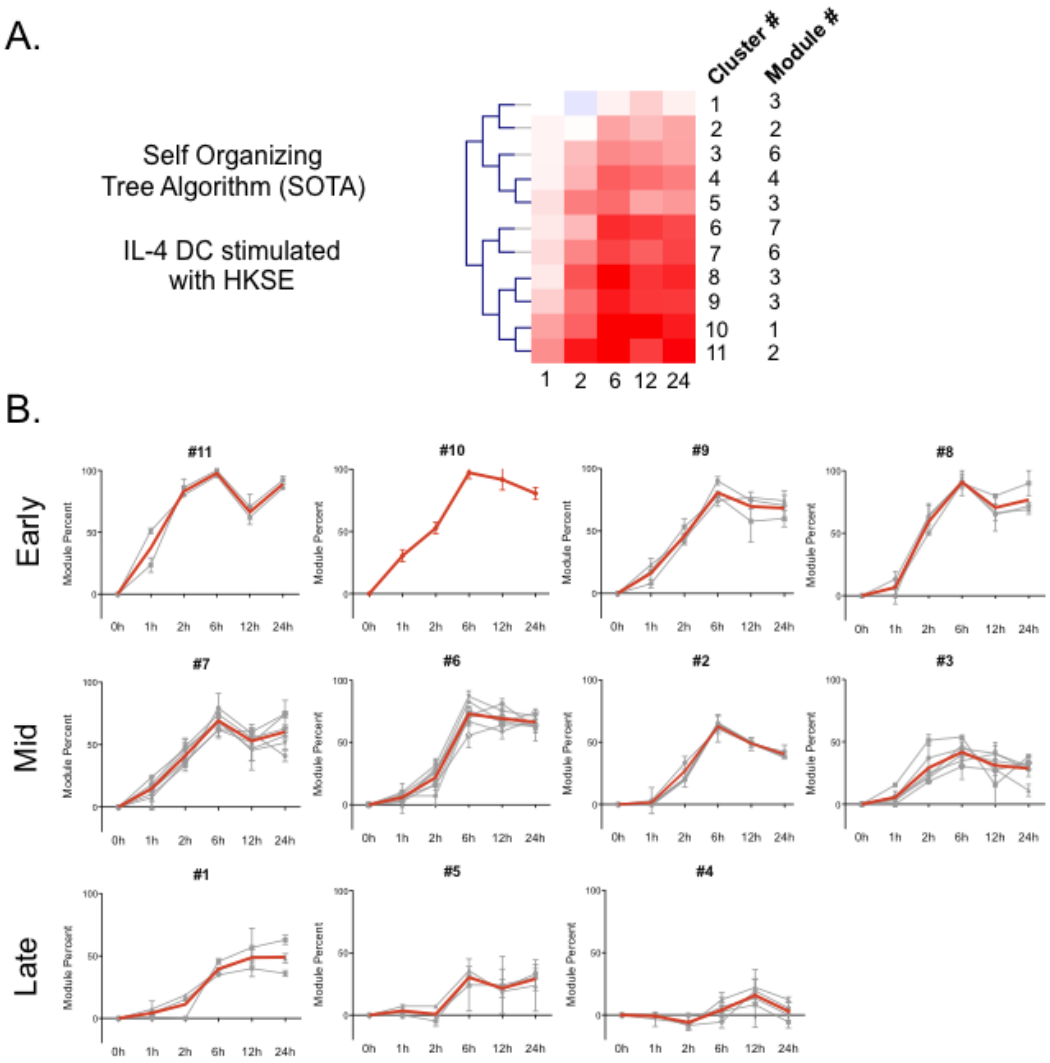


Figure 36: Kinetic profiles in HKSE-activated IL-4 DC

A. Self-organizing tree algorithm (SOTA) dendrogram representing the 11 longitudinal transcriptional profiles identified amongst modules over-expressed in at least one of the pathogen-activated DC time points. **B.** Line charts representing all modules (grey) for each profile. The average profile is represented in red.

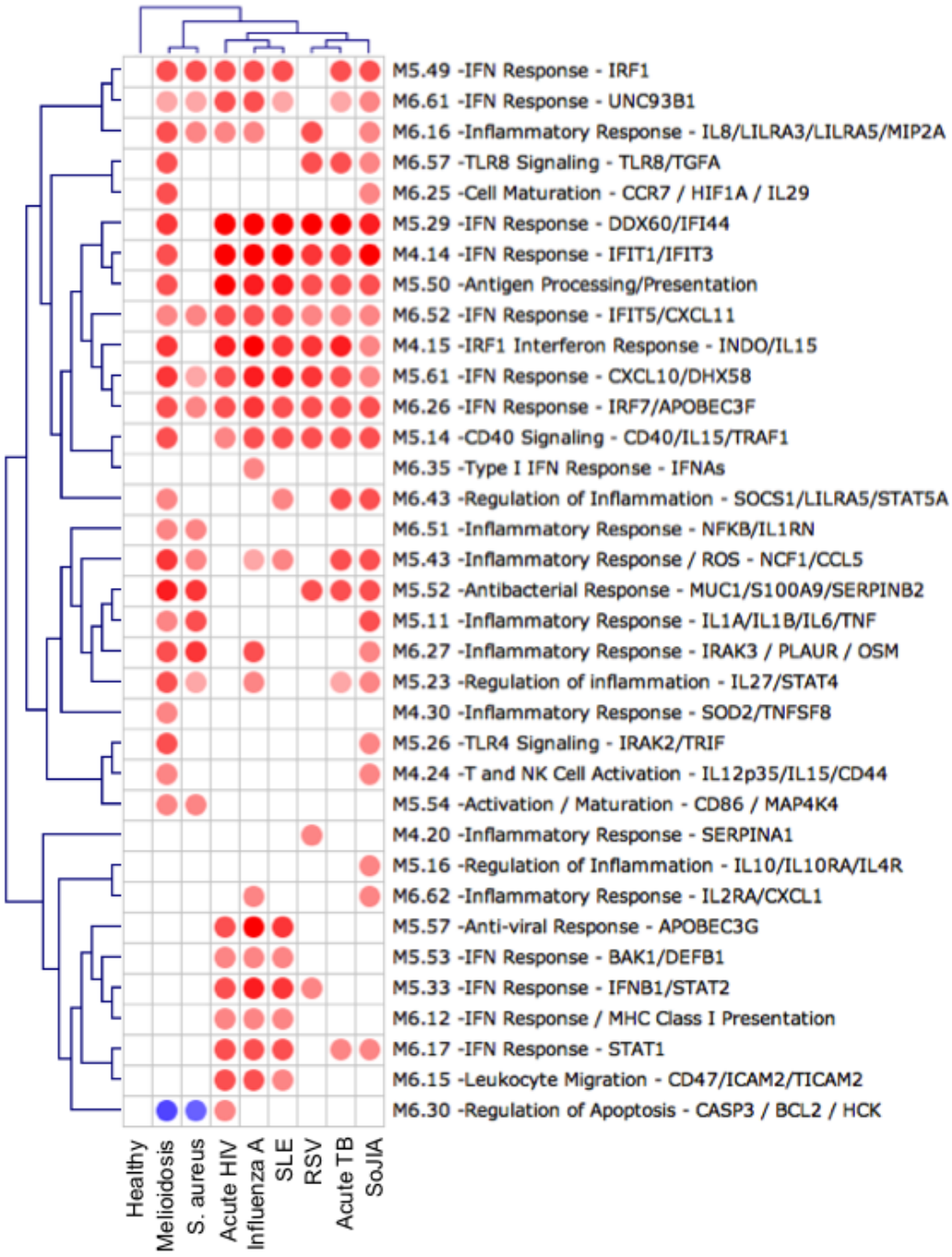


Figure 37: Application of DC modules to *ex vivo* whole blood datasets

Hierarchical clustering of DC modular fingerprints (Pearson) extracted from whole blood datasets from 6 infectious diseases (gram-negative sepsis melioidosis, community-acquired *S. aureus*, acute tuberculosis, influenza A, acute HIV, RSV) and 2 autoimmune diseases (SLE and SoJIA). Each disease group was normalized to its own healthy control group. The 35 modules displayed are over-expressed (>=30%) in at least one of the conditions.

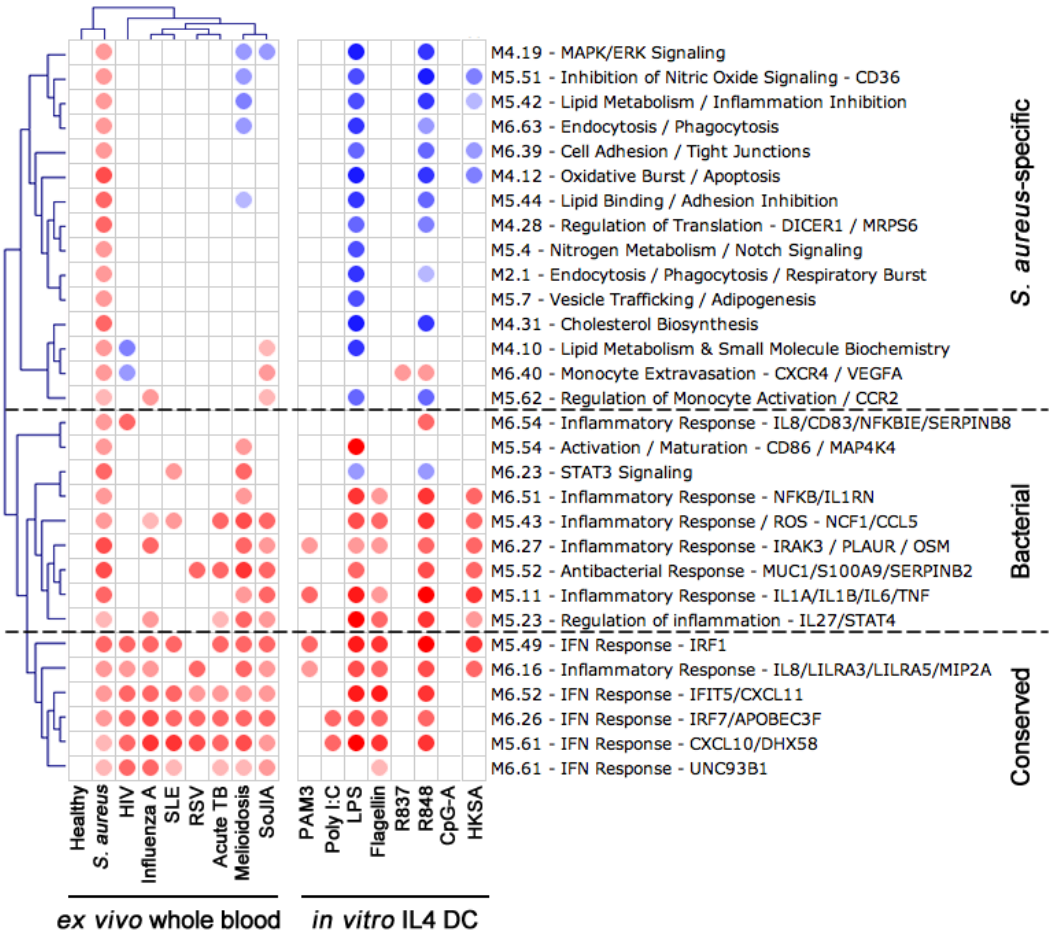


Figure 38: Lipid metabolism and endocytosis modules are over-expressed in patients with *S. aureus* infection

Hierarchical clustering (Pearson) of the 30 modules over-expressed (expression ≥ 30%)in the blood of patients with *S. aureus* infection and compared with the *ex vivo* whole blood signatures from other disease groups and in vitro signatures of IL4 DC stimulated for 6 hours with TLR ligands and heat-killed *S. aureus* (HKSA).

#	Dataset Name	Sample Count	Date	Cell Populations	Donors	Stimuli	Platform
1	TLR ligands, bacteria and cytokines	120	11/24/08	IFNa DC, IL4 DC	LB016 LB020 ND197	anti-CD40, CL097, cRPMI, Flagellin, HKEC, HKLM, HKPG, HKSA, IFNa, IL-10, IL-15, IL-1b, IL-21, LPS, PAM3, Poly I:C, R837, TNFa, TSLP, Zymosan	Illumina HT12v3
2	Flu Vaccine #1	24	12/22/09	IFNa DC, IL4 DC, IL15 DC	ND292	CL097, cRPMI, Flu Vaccine 6.25uL, Flu Vaccine 100uL, LPS, PAM3, Poly I:C, R837	Illumina HT12v3
3	Flu Vaccine #2	48	01/19/10	IFNa DC, IL4 DC, IL15 DC	ND284 ND294	CL097, cRPMI, Flu Vaccine 6.25uL, Flu Vaccine 100uL, LPS, PAM3, Poly I:C, R837	Illumina HT12v3
4	H1N1 Vaccine	22	03/16/10	IFNa DC, IL4 DC, IL15 DC	ND284 ND294	cRPMI, H1N1 Vaccine 6.25uL, LPS, Poly I:C	Illumina HT12v3
5	Live Viruses	95	05/04/10	IFNa DC, IL4 DC, IL15 DC	ND279 ND294	anti-CD40, cRPMI, Flu Media, H1N1 Brisbane, H1N1 Netherlands, H1N1 California, HIV JRCSF, HIV LAV, HIV Media, HKSA, LPS, PAM3, Poly I:C, PR8, R837, R848	Illumina HT12v3
6	TLR9, NOD2, RIG-I stimuli	44	12/02/10	IFNa DC, IL4 DC, IL15 DC	ND293 ND309 ND311	CpG 2006, CpG2216, cRPMI, MDP, Poly I:C-LMW-Lyovec	Illumina HT12v3
	Total	**353**					

*HKEC: heat-killed Escherichia coli; HKLM: heat-killed Listeria monocytogenes; HKPG: heat-killed Porphyromonas gingivalis; HKSA: heat-killed Staphylococcus aureus;

Table 19: Reference datasets used for the generation of antigen-presenting cell modules

Category	Ligand	Category	Source	Quantity	Target	IL4 DC	IFNa DC	IL15 DC
TLR Ligands & Cytoplasmic Receptors	PAM3-CSK4	Bacterial cell wall component	Invivogen	200ng/mL	TLR1/2	•	•	•
	Zymosan	Bacterial cell wall component	Invivogen	2ug/mL	TLR2/6	•	•	
	Poly I:C	Nucleic acid	Invivogen	10ug/mL	TLR3	•	•	•
	LPS	Bacterial cell wall component	Invivogen	100ng/mL	TLR4	•	•	•
	Flagellin	Bacterial cell wall component	Invivogen	100ng/mL	TLR5	•	•	
	R837	Nucleic acid	Invivogen	10ug/mL	TLR7	•	•	•
	R848	Nucleic acid	Invivogen	3ug/mL	TLR7/8	•	•	•
	CL097	Nucleic acid	Invivogen	5ug/mL	TLR7/8	•	•	•
	CpG 2006	Nucleic acid	Invivogen	1uM	TLR9	•	•	•
	CpG 2216	Nucleic acid	Invivogen	1uM	TLR9	•	•	•
	MDP	Bacterial cell wall component	Invivogen	10ug/mL	NOD2	•	•	•
	Poly I:C-LMW-Lyovec	Nucleic acid	Invivogen	1ug/mL	RIG-I	•	•	•
Cytokines	IL-1b	Cytokine	BIIR	100ng/mL	IL1R	•	•	
	IL-10	Cytokine	Peprotech	100ng/mL	IL10R	•	•	
	IL-15	Cytokine	Peprotech	100ng/mL	IL15R	•	•	
	IL-21	Cytokine	Invitrogen	100ng/mL	IL21R	•	•	
	TNFa	Cytokine	R&D	50ng/mL	TNFAR	•	•	
	IFNa	Cytokine	Schering-Plough	500U/mL	IFNAR	•	•	
	TSLP	Cytokine	BIIR	50ng/mL	?	•	•	
Heat-killed Bacteria	Heat-killed *E. coli*	Gram-negative bacteria	Invitrogen	10^8/mL	TLR4 / ?	•	•	
	Heat-killed *P. gingivalis*	Gram-negative bacteria	Invivogen	10^8/mL	TLR2 / ?	•	•	
	Heat-killed *L. monocytogenes*	Gram-positive bacteria	Invivogen	10^8/mL	TLR2 / ?	•	•	
	Heat-killed *S. aureus*	Gram-positive bacteria	Invivogen	10^8/mL	TLR2 / ?	•	•	
Viruses	Flu H1N1 Brisbane	Virus	BIIR	5:1 MOI	?	•	•	•
	Flu H1N1 California	Virus	BIIR	5:1 MOI	?	•	•	•
	Flu H1N1 Netherlands	Virus	BIIR	5:1 MOI	?	•	•	•
	PR8	Virus	BIIR	5:1 MOI	?	•	•	•
	HIV LAV	Virus	BIIR	1:1 MOI	?	•	•	•
	JRCSF	Virus	BIIR	1:1 MOI	?	•	•	•
	Flu Media	Control	BIIR		Control	•	•	•
	HIV Media	Control	BIIR		Control	•	•	•
Vaccines	Fluzone 2009 (6.25uL)	Vaccine	Sanofi Pasteur	6.25uL/500uL	?	•	•	•
	Fluzone 2009 (100uL)	Vaccine	Sanofi Pasteur	100uL/500uL	?	•	•	•
	H1N1 Vaccine 2009 (6.25uL)	Vaccine	Sanofi Pasteur	6.25uL/500uL	?	•	•	•
Other	anti-CD40	Other	BIIR	2ug/mL	CD40	•	•	•
	cRPMI	Control	Gibco		Control	•	•	•

• Samples present for at least 2 donors

Table 20: DC stimuli summary

Cluster ID	Modules	Annotation
#1	M6.8	T Cell Activation - CD70 / IL12B / CCL23 / HLA / ICAMs
	M6.25	Cell Maturation - CCR7 / HIF1A / IL29
	M6.27	Inflammatory Response - IRAK3 / PLAUR / OSM
#2	M4.30	Inflammatory Response - SOD2/TNFSF8
	M6.26	IFN Response - IRF7/APOBEC3F
#3	M5.33	IFN Response - IFNB1/STAT2
	M5.43	Inflammatory Response / ROS - NCF1/CCL5
	M5.54	Activation / Maturation - CD86 / MAP4K4
	M6.17	IFN Response - STAT1
	M6.35	Type I IFN Response - IFNAs
	M6.67	Granulocyte Chemottraction - CXCL6
#4	M5.50	Antigen Processing/Presentation
	M5.53	IFN Response - BAK1/DEFB1
	M6.30	Regulation of Apoptosis - CASP3 / BCL2 / HCK
	M6.61	IFN Response - UNC93B1
#5	M5.57	Anti-viral Response - APOBEC3G
	M6.56	Antibacterial Response - TNFRSF18/HAMP
	M6.57	TLR8 Signaling - TLR8/TGFA
#6	M3.4	T Cell Activation - CCL19/IL23A
	M4.20	Inflammatory Response - SERPINA1
	M5.16	Regulation of Inflammation - IL10/IL0RA/IL4R
	M5.23	Regulation of inflammation - IL27/STAT4
	M5.35	Inflammatory Response - NFKB1/IL10/CD93
	M5.52	Antibacterial Response - MUC1/S100A9/SERPINB2
	M6.43	Regulation of Inflammation - SOCS1/LILRA5/STAT5A
#7	M4.24	T and NK Cell Activation - IL12p35/IL15/CD44
	M5.14	CD40 Signaling - CD40/IL15/TRAF1
	M5.26	Undetermined - IRAK2/WTAP/F3
	M5.56	Anti-viral Response - IFIH1/BCL2A1/TRAF1
	M6.16	Inflammatory Response - IL8/LILRA3/LILRA5/MIP2A
	M6.28	IL1 signaling - NFKB2/PANX1/STARD5
#8	M5.29	IFN Response - DDX60/IFI44
	M5.61	IFN Response - CXCL10/DHX58
	M6.52	IFN Response - IFIT5/CXCL11
#9	M4.15	Inflammatory Response - INDO/IL15
	M6.51	Inflammatory Response - NFKB/IL1RN
	M6.62	Inflammatory Response - IL2RA/CXCL1
#10	M5.49	IFN Response - IRF1
#11	M4.14	IFN Response - IFIT1/IFIT3
	M5.11	Inflammatory Response - IL1A/IL1B/IL6/TNF

Table 21: DC time course SOTA results by cluster

Dataset	Pathogenesis	Patient Count	Healthy Control Count	Platform
Influenza A	Virus	21	14	Illumina HT12 V3
HIV	Virus	28	35	Illumina HT12 V3
RSV	Virus	52	14	Illumina HT12 V3
S. aureus	Gram+ Bacteria	99	44	Illumina HT12 V3
Burkholderia	Gram- Bacteria	18	5	Illumina HT12 V3
Acute Tuberculosis	Mycobacteria	23	11	Illumina HT12 V3
SLE	Autoimmune	55	14	Illumina HT12 V3
SoJIA	Autoimmune	62	23	Illumina HT12 V3

Table 22: Whole blood disease dataset summary

CHAPTER 6: DISCUSSION & CONCLUSIONS

Summary

The work presented herein relied on the use of systems biology approaches to study the status of the host immune system in response to *S. aureus* infection. To this end, the transcriptional profiles of PBMC and whole blood from patients with community-acquired *S. aureus* infection were characterized by microarray and and leukocyte population frequencies assessed by polychromatic flow cytometry. Furthermore, an *in vitro* system of antigen-presenting cell stimulation with various pathogens, including *S. aureus* as well as other bacteria and viruses, was used to identify early inflammatory programs induced in innate immune cells in response to infection. As these high-throughput comprehensive approaches generate large amounts of data, I relied on a collection of modular frameworks to analyze and interpret the *ex vivo* and *in vitro* fingerprints obtained. This approach first led us to identify the global signature of *S. aureus* infection in PBMC and correlate it with frequencies of circulating leukocyte populations. I then expanded these observations in whole blood, focusing on transcriptional correlates of clinical heterogeneity. Finally, I developed an *in vitro* APC analytical framework based on the response of monocyte-derived DC to a variety of pathogens and simpler stimuli, to identify the different components of the inflammatory response in infection.

Transcriptional profiles of PBMC from patients with S. aureus infection

The versatility of *S. aureus* has allowed emergence of highly resistant and virulent bacterial strains in the community[172]. Its pathogenicity is in part due to its repertoire of virulence factors and proclivity for tissue and endovascular invasion, destruction, and dissemination while simultaneously evading multiple components of the innate immune system and secreting immunomodulatory proteins that compromise both humoral and cell-mediated immunity[172,192,193]. Little is known about the relationship between the human host and *S. aureus* during invasive infections, which lead us to carry out the studies presented herein. We applied a systems biology approach utilizing both gene expression microarray profiling and corresponding flow cytometry analyses to assess the host immune status during infection.

By conducting several different step-wise analyses, the gene expression profile of patients with invasive *S. aureus* infections was defined and its robustness validated among distinct patient populations and across microarray platforms. Modular analysis demonstrated significant activation of host genes related to the innate immune response with increased expression of genes related to inflammatory processes and cells of the myeloid lineage and significant under-expression of genes related to the adaptive immune response. Although the over-expression of the innate immune response genes was not unexpected[193,194], the striking and consistent decreased expression of B and T cell-related genes was less anticipated.

There is limited and conflicting information regarding the numbers of lymphocyte populations in patients with acute *S. aureus* infections[195-197]. In one study,

patients with *S. aureus* and *S. pneumoniae* sepsis showed significantly decreased numbers of CD4+ and CD8+ T cells, and NK cells[198], while in another, patients with MRSA superantigen-associated glomerulonephritis showed increased numbers of DR+ CD4+ and CD8+ T cells and NK cells[198]. Flow cytometry analyses were performed in our study subjects to better understand whether the changes observed in the gene expression patterns simply reflected alterations in immune cell numbers. Despite significant under-expression of T and B cell-related genes observed in our patients with *S. aureus* infections, there were no differences in the absolute numbers of total B and T cells between infected patients and healthy controls in the PBMC study. At that time no consistent significant differences were seen between patients and controls across B cell subpopulations evaluated, however possible trends in transitional and pre-germinal B cells prompted us to increase our sample size. With subsequent addition of 14 patients, we detected a small yet significant increase in all circulating B cell populations (including transitional, naïve, and memory B cells), with the exception of plasma cells.

Detailed analysis of the T cell sub-populations revealed decreased numbers of both central memory CD4+ T cells and CD8+ T cells, but no difference in the other T cell subsets in patients with *S. aureus* infection. Central memory T cells have been shown to have a high proliferative potential and to demonstrate *in vivo* persistence[199,200]. Our results demonstrated a decreased number of central memory T cells, suggesting a possible reorganization of the circulating T cell compartment that could also explain the reduced expression of T cell-related genes. With increased expression of CCR7 and CD62L, central memory T cells are programmed to preferentially migrate to lymphoid tissues to interact with other T and B cells in establishing a repertoire of effector

functions against the invading pathogen[186]. One possible explanation then, for the significant reduction in the numbers of circulating central memory T cells seen in our subjects with *S. aureus* infection may be central memory T cell homing to these secondary lymph organs. *S. aureus* expresses factors that promote its immune evasion and could also elicit this cellular imbalance. *In vitro* observations have demonstrated a shift from central memory to effector memory T cells in the presence of *S. aureus* enterotoxin[186]. Secretion of other superantigens by *S. aureus*, such as Map/Eap (MHC Class II analogous protein/extracellular adherence protein), bind to T cell receptors and directly thwart T cell responses by reducing T cell proliferation, altering effector functions, and stimulating apoptosis[192].

Alternatively, the under-expression of T cell genes may be due to an increase in other cell populations. Indeed, the monocyte compartment was expanded in average by 4-fold in patients with *S. aureus* infection and correlated with gene expression levels, providing a cellular correlate to the increased expression of innate immune genes in PBMC. Based on CD16 expression, circulating monocytes can be divided into functionally distinct subpopulations CD14+CD16- and CD14+16+ monocytes, which may have distinct roles in the innate immune response[201]. In *S. aureus* patients, there was a significant expansion of both monocyte subsets. Furthermore, the high levels of expression of genes generally associated with neutrophil functions, such as defensins, lactotransferrins, CAMP, and elastase, as seen in PBMC module M2.2, may originate from the CD14+16+ (CD62L-) monocytes subset, as has been previously described[201]. Although significant numbers of neutrophils are not generally present in PBMCs, low-density neutrophils have been demonstrated and accounted for over-expression of

neutrophil genes in patients with SLE[116]. Despite using similar technical approaches, we were not able to detect neither these low-density immature neutrophils nor high-density neutrophils that could result from contamination in our samples. Studies with purified cell subsets will be necessary to understand the distinct transcriptional contribution from each population.

Transcriptional profiles of whole blood from patients with S. aureus infection

While the PBMC study allowed us to identify a signature of *S. aureus* infection common to all patients, it has several shortcomings. First, the transcriptional profiling of PBMC requires real time manipulation of fresh blood samples and is therefore not practical for large-scale studies and clinical applications. Second, cells depleted from PBMC such as neutrophils, which play a major role in the defense against *S. aureus*, might provide important information for the systemic characterization of these infections[187]. Finally, our initial study focused on defining molecular profiles common to all patients, and did not address signature heterogeneity in the context of clinical disease variability.

To address these issues, we conducted a new study of whole blood transcriptional patterns in a new cohort of 99 pediatric patients with acute community-acquired *S. aureus* infections. The heterogeneity of clinical presentation, the acute nature of these infections, the complex treatment regimen and the delayed characterization of the pathogen rendered the study design and analysis of patient immune status during staphylococcal infections challenging. Using whole blood transcriptional profiling, we have identified both

conserved and heterogeneous elements of the signature that correlate with time elapsed since hospital entry, dissemination and clinical presentation.

Supporting the observations from the PBMC study, the whole blood signature was characterized by significant over-expression of transcripts linked to the myeloid lineage and inflammation, and under-expression of transcripts linked to the lymphoid lineage. This reflects the large expansion of circulating neutrophils and monocytes during the acute phase of infection as detected in patient's routine hemograms. Additionally, a family of transcripts linked to hematopoiesis and erythropoiesis were over-expressed is a subgroup of patients, perhaps due to an increased release of hematopoietic precursors from the bone marrow during acute infection[202]. We did not observe consistent changes in transcripts linked to the interferon response, suggesting that circulating leukocytes do not mediate response to *S. aureus* infection through this pathway, as was observed during active pulmonary tuberculosis[133]. It is however noticeable that 15 out of 99 patients displayed over-expression of the three IFN modules. All these patients but one tested negative for concomitant viral infections. While it is possible that this represents a signature of recently cleared viral infection, it could also reflect an activation of the IFN pathway to counter arrest the bacteria-induced pro-inflammatory milieu[203].

Currently, there is no established laboratory marker to objectively define the clinical disease severity in patients with *S. aureus* infections or to monitor the spread of the infection to multiple organs[167]. Transcriptional profiling highlighted global quantitative differences between patients with local or disseminated disease, supporting the potential role of microarrays and the associated quantitative MDTH in monitoring the spread of infection. We were also able to connect/correlate "molecular nodes" (modules)

with “clinical nodes” (laboratory parameters), suggesting that the patient’s transcriptional score for one or several modules could be used as an additional measure of hematopoiesis and systemic immune status during acute bacterial infections. Future studies should focus on the predictive value of these approaches in identifying patients at risk of infection dissemination, to adapt treatment accordingly.

The heterogeneity of clinical presentation led us to analyze signatures on an individual basis. We approached this analysis from two angles. The traditional approach consisted in grouping patients according to clinical diagnoses, based on a classification defined by an independent clinician, and conduct statistical tests between clinically-defined groups of patients. The other approach consisted in identifying groups of patients based on molecular profiles, without a priori knowledge of clinical classification. Both approaches identified groups (or clusters) of patients with various inflammation, blood cell distribution and infection stage. Molecular heterogeneity was observed both quantitatively (MDTH measurements) and qualitatively (clustering of patients and modules). The major transcriptional patterns identified included a pro-inflammatory myeloid signature, linked to blood draw early in the course of the diseases, high circulating neutrophil and monocyte counts and elevated CRP (C1). A second transcriptional pattern was characterized by an erythropoiesis signature with reduced myeloid components (C3). This erythropoiesis signature was observed by others in systemic onset juvenile idiopathic arthritis (SOJIA)[190] and was proposed to reflect the expansion of immature precursor cells and ineffective erythropoiesis. The latter results in accumulation of iron[204] in tissues, which has implications for bacterial survival. Alternatively, increased erythropoiesis may reflect increased tissue hypoxia in this group

of patients. Interestingly, a recent study suggested that erythropoietin (EPO) inhibits NF-kB and TNFa-mediated pro-inflammatory pathways[205]. This observation could explain why patients from C3 displayed limited pro-inflammatory signature. Again, this could represent a bacterial survival mechanism, whereby increased erythropoiesis would prevent the development of an adequate pro-inflammatory response and subsequent bacterial clearance.

Finally, we compared two distinct clinical presentations, pneumonia and osteoarticular infections, and found that patients with osteoarticular infections displayed over-expression of transcripts linked to blood coagulation. This supports previous studies demonstrating increased systemic coagulation in patients with musculoskeletal infections and their increased risk of developing deep venous thrombosis (DVT) and more severe pulmonary emboli[206-208]. Blood microarray profiling could thus serve as an additional readout of coagulation state and inform physicians about potential need for anti-coagulant treatment. Additionally patients with pneumonia displayed over-expression of genes involved in cholesterol synthesis, which has recently been described to be involved in neutrophil recruitment to the lung[209]. While it is difficult to differentiate an active process from a side-effect of immune activation, tracking cholesterol-related transcriptional programs in whole blood during acute pulmonary infection may be of interest to infectious disease clinicians as a potential biomarker to track systemic inflammation onset and resolution.

Patients were recruited at different stages of hospitalization and the acute, intense nature of these infections may result in major changes in blood transcriptional signature over the course of infection. We addressed this issue by assessing how draw index and

disease onset index influenced the intensity of the signature. It confirmed that global signature intensity decreases as patients get closer to discharge independently of the localization of the infection or clinical presentation, suggesting that microrarray profiling reflects well the evolution of patient status during acute bacterial infections. Further longitudinal studies will be required to assess the predictive value of blood signatures regarding disease course and and clinical heterogeneity. Overall, blood transcriptional profiling stands as an all-in-one additional tool for clinicians and scientists to analyze the qualitative and quantitative systemic status of the host immune response/ immune system of patients with acute *S. aureus* infection.

Transcriptional profiles of antigen-presenting cells activated by pathogen in vitro

The *ex vivo* studies in PBMC and whole blood highlighted the induction of innate inflammatory pathways in the leukocytes of patients with *S. aureus* infection. Increased frequencies of antigen-presenting cells, including mDC, pDC and most circulating B cell populations measured by flow cytometry during the PBMC study prompted us to develop an *in vitro* model of APC stimulation with *S. aureus* and other pathogens. To do so, we generated dendritic cells *in vitro*, derived from healthy donor monocytes cultured in the presence of GM-CSF and either of three cytokines (IL-4, IFNa and IL-15), to mimic various inflammatory conditions. These DC were subsequently stimulated for 6 hours with a variety of artificial ligands or whole pathogens targeting an array of innate immune receptors, and transcriptional profiles were measured by microarray. To analyze the large amount of data generated this way, I leveraged the module framework

technology used in both PBMC and whole blood studies to develop a new set of inflammatory DC modules. I annotated the modules using a combination of knowledge-driven and data driven approaches. I then applied this framework to compare the responses of IL-4 and IFNa DC to flu (H1N1 Brisbane), *Salmonella enterica* or *Staphylococcus aureus in vitro*. Finally, I applied this framework to *ex vivo* whole blood fingerprints from patients with various infectious or autoimmune diseases to further delineate inflammatory networks involved.

The extension of *ex vivo* whole blood analysis into an *in vitro* model of antigen-presenting cell activation provided several additional benefits. The modular frameworks developed for PBMC and whole blood contained modules that represented very distinct leukocyte populations (T cells, plasma cells, myeloid cells, NK cells), or broad immune mechanisms (inflammation, IFN response), without much precision on the specific pathways involved. Interpretation of circulating leukocyte fingerprints with these frameworks resulted in signatures highly influenced by changes in cell population frequencies, as could be observed in both *ex vivo S. aureus* studies, with myeloid and inflammatory modules over-expressed. The development of modules from relatively homogeneous monocyte-derived DC populations stimulated *in vitro* resulted on the other hand in a collection of modules focused on dynamic cell-intrinsic innate immunity pathways, which provided additional breakdown of inflammatory and interferon responses.

The DC framework is an analytical tool with an array of potential applications, including: i) the enrichment of molecules involved in innate responses through module connectivity; ii) the characterization of pathogen components signaling pathways; iii) the

phenotyping of APC subsets at baseline or in response to *in vitro* stimulation; iv) the characterization of disease inflammatory components in circulating leukocytes. The application of this framework to whole blood transcriptional profiles from infectious and autoimmune diseases identified clear DC module fingerprints. It is important to understand that these profiles cannot at this stage be attributed to a specific cell population in the blood, such as myeloid or plasmacytoid DC. The modules were obtained from monocyte-derived DC profiles, so it is likely that most circulating myeloid cells, including monocytes, blood DC but also neutrophils, could display changes in DC module fingerprints. It was previously observed that the interferon signature detected in the blood of patients with acute tuberculosis (which is also detected with the DC modules) mostly comes from neutrophils[133]. Future studies should explore the behavior of these transcriptional modules in sorted cell populations, in healthy and disease states.

To build the DC modules, I chose 6-hour as the reference time point, because it is late enough to detect the bulk of primary transcriptional responses, and early enough to avoid the confounding secondary transcriptional responses arising from positive and negative feedback loops. Huang et al[105], who previously performed *in vitro* DC transcriptional kinetics, distinguished three waves of transcription (early, middle and late phases). Each wave is additionally split in genes that are either sustained or transient. By 6 hours, only the early phase has reached its peak. The middle and late phase peak out around 12 and 18 hours respectively. Developing the modules using a 6-hour reference time point may thus have filtered out transcripts that are either induced early on and transiently (between 0 and 4h), as well as all transcripts appearing later than 6 hours. It is therefore important to understand the context of development of these frameworks and

consider non-selected transcripts during analysis. Future work should focus on developing frameworks that take into account kinetics of transcription.

Adapting an analytical framework developed from DC generated *in vitro* from human monocytes to the study of inflammatory components of whole blood transcriptional profiles may seem counterintuitive at first. As mDC and pDC are present at low frequencies in whole blood, their contribution to the global blood signature will under most conditions be masked by larger populations including monocytes, neutrophils or T cells. There is however growing evidence that under certain *in vivo* inflammatory conditions, monocytes can mature into antigen-presenting cells with DC phenotype[210,211]. Furthermore, healthy monocytes cultured with serum from SLE, a disease with a strong type I interferon component, were previously shown to mature into DC[212]. In this context, the maturation of monocyte *in vitro* in the presence of IFNa or IL-4 can be likened to *in vivo* antiviral or SLE-like and Th2 like conditions respectively. The *in vitro* DC modules may thus serve as a good readout of *in vivo* monocyte activation during various inflammatory conditions.

In this context, the application of the DC modules to whole blood datasets from patients with infectious and autoimmune diseases helped us further decorticate the inflammatory programs involved in pathogenesis. We identified subgroups of IFN responses, with different disease specificities. Additionally, we identified a group of modules linked to lipid metabolism and cellular compartmentalization that was specific to *S. aureus* patients, and was not part of the 44 modules over-expressed by some or most *in vitro* innate stimuli. This supports the observed outcome of *in vitro* infection of adipocyte-like cell lines by *S. aureus*, which inhibits cellular lipidosis and induces the

release of inflammatory adipokines by infected cells[213]. Interestingly, this transcriptional program is absent in patients with melioidosis caused by the gam-negative bacteria *B. mallei*. We cannot exclude that this inflammatory program is common to gram-positive bacteria, which warrants the further examination of whole blood transcriptional profiles in other gram-positive infections such as listeriosis, or Streptococcal pneumonia and meningitis.

Perspectives

The work presented herein relies on the use of microarrays to measure transcriptional changes in blood leukocytes to better characterize the host immune response to *S. aureus* infection. We identified a common signature of infection, both in PBMC and whole blood, as well as heterogeneous transcriptional programs that correlated with the time of blood sampling, the extent of dissemination and the site of infection (joint versus lung). As the collection of longitudinal samples in pediatric cohorts in these acute bacterial infections is challenging, the analysis was conducted cross-sectionally. All samples were collected after patients were initially treated, which raises questions about the effect of treatment on transcriptional fingerprint obtained. Future studies should aim at collecting blood longitudinally, starting at the time of admission and until discharge. Analysis of follow-up samples after discharge may also be of interest, as the signature was not extinguished in all patients on the day of discharge.

One characteristic of *S. aureus* infections is their capacity to start locally, under the form of skin abscesses, and disseminate through the blood stream to cause life-

threatening disease. Because our study design did not permit it, we did not assess the capacity of microarrays to predict dissemination of infection, which may be of interest to clinicians in the treatment of patients. Furthermore, the sole clinical outcome in our studies was recovery and discharge, so we were not able to associate signatures with other (worse) outcomes. Future studies should address the capacity of blood microarrays to determine length of stay and clinical outcomes based on the type and frequency of treatment dispensed.

While the microarray technology will probably not become a generic assay in clinical practice, as it still requires extensive time between data collection and interpretation, and as it is quickly being replaced by more comprehensive assays such as sequencing, which provides absolute mRNA quantitation as well as additional genetic and epigenetic information, I believe it provides a good proof-of-principle for these all-in-one assays. The dimension and complexity of the data generated by these technologies requires systems biologists to develop tools that facilitate their interpretation by people that are not specially trained for it. The modular frameworks presented herein are an attempt to answer this need, by providing a simple snapshot of the status of the immune system during disease, or *in vitro* stimulation. Additionally, these fingerprints can be complemented by molecular scores such as the MDTH that informs the analyst about the extent of transcriptional perturbations on an individual basis. From global disease fingerprints, we now need to better understand how these analytical tools can explain disease heterogeneity and further understand pathogenesis. Comparing *ex vivo* profiles to *in vitro* stimulation assays may help identify signaling pathways involved as well as the nature of the stimuli responsible for onset and continuation of disease. Further studies

will be required to better characterize the modules obtained, including their major upstream regulators and their functional significance at the systemic level. This will lead to a better understanding of mechanisms involved in disease pathogenesis, a faster identification of biomarkers and therapeutic targets, and hopefully more effective treatments for immune-mediated diseases.

Appendix A – Functional interpretation of PBMC transcriptional modules

Module I.D.	Probes Count	Keyword Selection	Assessment
M 1.1	76	Ig, Immunoglobulin, Bone, Marrow, PreB, IgM,Mu.	**Plasma cells.** Includes genes coding for Immunoglobulin chains (*e.g.* IGHM, IGJ, IGLL1, IGKC, IGHD) and the plasma cell marker CD38.
M 1.2	130	Platelet, Adhesion, Aggregation, Endothelial, Vascular	**Platelets.** Includes genes coding for platelet glycoproteins (ITGA2B, ITGB3, GP6, GP1A/B), and platelet-derived immune mediators such as PPPB (pro-platelet basic protein) and PF4 (platelet factor 4).
M 1.3	80	Immunoreceptor , BCR, B-cell, IgG	**B-cells.** Includes genes coding for B-cell surface markers (CD72, CD79A/B, CD19, CD22) and other B-cell associated molecules: Early B-cell factor (EBF), B-cell linker (BLNK) and B lymphoid tyrosine kinase (BLK).
M 1.4	132	Replication, Repression, Repair, CREB, Lymphoid, TNF-alpha	**Undetermined.** This set includes regulators and targets of cAMP signaling pathway (JUND, ATF4, CREM, PDE4, NR4A2, VIL2), as well as repressors of TNF-alpha mediated NF-kappa B activation (CYLD, ASK, TNFAIP3).
M 1.5	142	Monocytes, Dendritic, MHC, Costimulatory, TLR4, MYD88	**Myeloid lineage.** Includes molecules expressed by cells of the myeloid lineage (CD86, CD163, FCGR2A), some of which being involved in pathogen recognition (CD14, TLR2, MYD88). This set also includes TNF family members (TNFR2, BAFF).
M 1.6	141	Zinc, Finger, P53, RAS	**Undetermined.** This set includes genes coding for signaling molecules, *e.g.* the zinc finger containing inhibitor of activated STAT (PIAS1 and PIAS2), or the nuclear factor of activated T-cells NFATC3.
M 1.7	129	Ribosome, Translational, 40S, 60S, HLA	**MHC/Ribosomal proteins.** Almost exclusively formed by genes coding MHC class I molecules (HLA-A,B,C,G,E)+ Beta 2-microglobulin (B2M) or Ribosomal proteins (RPLs, RPSs).
M 1.8	154	Metabolism, Biosynthesis, Replication, Helicase	**Undetermined.** Includes genes encoding metabolic enzymes (GLS, NSF1, NAT1) and factors involved in DNA replication (PURA, TERF2, EIF2S1).
M 2.1	95	NK, Killer, Cytolytic, CD8, Cell-mediated, T-cell, CTL, IFN-g	**Cytotoxic cells.** Includes cytotoxic T-cells amd NK-cells surface markers (CD8A, CD2, CD160, NKG7, KLRs), cytolytic molecules (granzyme, perforin, granulysin), chemokines (CCL5, XCL1) and CTL/NK-cell associated molecules (CTSW).
M 2.2	49	Granulocytes, Neutrophils, Defense, Myeloid, Marrow	**Neutrophils.** This set includes innate molecules that are found in neutrophil granules (Lactotransferrin: LTF, defensin: DEAF1, Bacterial Permeability Increasing protein: BPI, Cathelicidin antimicrobial protein: CAMP).

M 2.3	148	Erythrocytes, Red, Anemia, Globin, Hemoglobin	**Erythrocytes.** Includes hemoglobin genes (HGBs) and other erythrocyte-associated genes (erythrocytic alkirin:ANK1, Glycophorin C: GYPC, hydroxymethylbilane synthase: HMBS, erythroid associated factor: ERAF).
M 2.4	133	Ribonucleoprotei n, 60S, nucleolus, Assembly, Elongation	**Ribosomal proteins.** Including genes encoding ribosomal proteins (RPLs, RPSs), Eukaryotic Translation Elongation factor family members (EEFs) and Nucleolar proteins (NPM1, NOAL2, NAP1L1).
M 2.5	315	Adenoma, Interstitial, Mesenchyme, Dendrite, Motor	**Undetermined.** This module includes genes encoding immune-related (CD40, CD80, CXCL12, IFNA5, IL4R) as well as cytoskeleton-related molecules (Myosin, Dedicator of Cytokenesis, Syndecan 2, Plexin C1, Distrobrevin).
M 2.6	165	Granulocytes, Monocytes, Myeloid, ERK, Necrosis	**Myeloid lineage.** Includes genes expressed in myeloid lineage cells (IGTB2/CD18, Lymphotoxin beta receptor, Myeloid related proteins 8/14 Formyl peptide receptor 1), such as Monocytes and Neutrophils.
M 2.7	71	No keywords extracted.	**Undetermined.** This module is largely composed of transcripts with no known function. Only 20 genes associated with literature, including a member of the chemokine-like factor superfamily (CKLFSF8).
M 2.8	141	Lymphoma, T-cell, CD4, CD8, TCR, Thymus, Lymphoid, IL2	**T-cells.** Includes T-cell surface markers (CD5, CD6, CD7, CD26, CD28, CD96) and molecules expressed by lymphoid lineage cells (lymphotoxin beta, IL2-inducible T-cell kinase, TCF7, T-cell differentiation protein mal, GATA3, STAT5B).
M 2.9	159	ERK, Transactivation, Cytoskeletal, MAPK, JNK	**Undetermined.** Includes genes encoding molecules that associate to the cytoskeleton (Actin related protein 2/3, MAPK1, MAP3K1, RAB5A). Also present are T-cell expressed genes (FAS, ITGA4/CD49D, ZNF1A1).
M 2.10	106	Myeloid, Macrophage, Dendritic, Inflammatory, Interleukin	**Undetermined.** Includes genes encoding for Immune-related cell surface molecules (CD36, CD86, LILRB), cytokines (IL15) and molecules involved in signaling pathways (FYB, TICAM2-Toll-like receptor pathway).
M 2.11	176	Replication, Repress, RAS, Autophosphoryla tion, Oncogenic	**Undetermined.** Includes kinases (UHMK1, CSNK1G1, CDK6, WNK1, TAOK1, CALM2, PRKCI, ITPKB, SRPK2, STK17B, DYRK2, PIK3R1, STK4, CLK4, PKN2) and RAS family members (G3BP, RAB14, RASA2, RAP2A, KRAS).
M 3.1	122	ISRE, Influenza, Antiviral, IFN-gamma, IFN-alpha, Interferon	**Interferon-inducible.** This set includes interferon-inducible genes: antiviral molecules (OAS1/2/3/L, GBP1, G1P2, EIF2AK2/PKR, MX1, PML), chemokines (CXCL10/IP-10), signaling molecules (STAT1, STAt2, IRF7, ISGF3G).
M 3.2	322	TGF-beta, TNF, Inflammatory, Apoptotic, Lipopolysacchari de	**Inflammation I.** Includes genes encoding molecules involved in inflammatory processes (*e.g.* IL8, ICAM1, C5R1, CD44, PLAUR, IL1A, CXCL16), and regulators of apoptosis (MCL1, FOXO3A, RARA, BCL3/6/2A1, GADD45B).

M 3.3	276	Inflammatory, Defense, Lysosomal, Oxidative, LPS	**Inflammation II.** Includes molecules inducing or inducible by inflammation (IL18, ALOX5, ANPEP, AOAH, HMOX1, SERPINB1), as well as lysosomal enzymes (PPT1, CTSB/S, NEU1, ASAH1, LAMP2, CAST).
M 3.4	325	Ligase, Kinase, KIP1, Ubiquitin, Chaperone	**Undetermined.** Includes protein phosphatases (PPP1R12A, PTPRC, PPP1CB, PPM1B) and phosphoinositide 3kinase (PI3K) family members (PIK3CA, PIK32A, PIP5K3).
M 3.5	22	No keyword extracted	**Undetermined.** Composed of only a small number of transcripts. Includes hemoglobin genes (HBA1, HBA2, HBB).
M 3.6	288	Ribosomal, T-cell, Beta-catenin	**Undetermined.** This set includes mitochondrial ribosomal proteins (MRPLs, MRPs), mitochondrial elongations factors (GFM1/2), Sortin Nexins (SN1/6/14) as well as lysosomal ATPases (ATP6V1C/D).
M 3.7	301	Spliceosome, Methylation, Ubiquitin	**Undetermined.** Includes genes encoding proteasome subunits (PSMA2/5, PSMB5/8); ubiquitin protein ligases HIP2, STUB1, as well as components of ubiqutin ligase complexes (SUGT1).
M 3.8	284	CDC, TCR, CREB, Glycosylase	**Undetermined.** Includes genes encoding enzymes: aminomethyltransferase, arginyltransferase, asparagines synthetase, diacylglycerol kinase, inositol phosphatases, methyltransferases, helicases…
M 3.9	260	Chromatin, Checkpoint, Replication, Transactivation	**Undetermined.** Includes genes encoding kinases (IBTK, PRKRIR, PRKDC, PRKCI) and phosphatases (*e.g*. PTPLB, PPP2CB/3CB, PTPRC, MTM1, MTMR2).

Appendix B – Functional interpretation of whole blood transcriptional modules

Module I.D.	Probes Count	Annotation	Key Transcripts
M1.1	92	Coagulation / Platelets	F13A1, CLEC1B, ALOX12, VWF, PPBP, PF4V1, ITGA2B
M1.2	36	Interferon Response	CXCL10, DDX60, IFI44, IFI44L, IFIT1, IFIT3, IFITM3, OAS1, OAS2, OAS3, OASL, OTOF
M2.3	75	Erythrocyte Development	ALAS2, ERAF, GYPB, GYPE, HBD, RHAG
M3.1	89	Erythrocyte Development	GYPC, HBG1, HBG2, HBQ1, KEL, VWCE, SPTA1
M3.2	148	Myeloid Lineage	TLR4, TLR6, TLR8, SERPINA1, SIGLEC5, S100A9, IL1RAP, IL1RN, ITGAM, LILRA3, F5. CEACAM4, AQP9
M3.3	49	Hematopoiesis / Cell Cycle	CDC2, CDC20, CDCA5, CKS2
M3.4	62	Interferon Response	DDX58, DHX58, IFI35, IFIT2, IFIT5, IRF7, INDO, OAS2, STAT1, STAT2, GBP1, GBP3-6
M3.5	149	Protein Synthesis	Transcription factors, translation factors, ubiquitination, ribosomal proteins
M3.6	54	Cytotoxicity / NK Cells	CD8A, GNLY, GZMA, GZMH< GZMM, IFNG, KIR3DL2, KLRC3, KLRD1, NCR3, NKG7, PRF1
M4.1	68	T Cells	CD3E, CD28, CD40LG, IL4R, ICOS, TCF7, TRAT1
M4.2	52	Inflammation	IRAK3, IL1R2, IL18R1, IL18RAP, CASP5, OSM, PGLYRP1, OPLAH, MMP9, TLR5
M4.3	74	Protein Synthesis	Ribosomal proteins and predicted ribosomal proteins
M4.4	78	Erythrocyte Development	ERMAP, GYPC
M4.5	86	Protein Synthesis	Ribosomal proteins, translation elongation factors, ubiquitination
M4.6	116	Myeloid Lineage	STAT3, STAT5B, CLIC1, CKLF, CD97
M4.7	98	Lymphoid Lineage	Molecules enriched in T cell and B cells according to BioGPS
M4.10	35	B Cells	CD19, CD79A, CD79B, BANK1, BLK, EBF1, VPREB3
M4.11	20	Plasma Cells	CD38, IGJ, TNFRSF17
M4.13	82	Myeloid Lineage	IL13RA1, IL1B, IL8RA, IL8RB, P2RY13, NCF4, TLR6
M4.14	62	Myeloid Lineage	CD33, CD1D, CD36, CD86
M4.15	45	Cytotoxicity / NK Cells	CD2, CD27, ITK, LCK, LY9, PRKCH, PRKCQ
M5.1	244	Inflammation	GRN, DUSP18, SERPINB8, TOLLIP
M5.3	101	Hematopoiesis	EIFs, JUND, ubiquitination
M5.9	84	Protein Synthesis	Ribosomal proteins
M5.12	63	Interferon Response	IFI16, IRF9, TAP1, TAP2, SP100, SP110, SP140, TRIM5, TRIM25, TRIM38, TRIM56, UNC93B1
M5.15	24	Neutrophils	AZU1, BPI, CAMP, CEACAM6, CEACAM8, CTSG, DEFA4, ELA2, LTF, MPO, RETN, OLR1
M6.6	53	Myeloid Lineage	CD14, CD300C, IL4R, CTSD

M6.7	76	Lymphoid Lineage	CD47, transcripts enriched in B and T cells
M6.9	38	Lymphoid Lineage	LAX1, ribosomal proteins
M6.11	18	Cell Cycle	CDC25A, CDCA3, CENPE, KIFs, NUF2
M6.12	70	Protein Synthesis	Translation factors, ribosomal proteins
M6.13	47	Inflammation	LILRA2, LILRB3, LAMP2, FCER1G, PILRA, PLAUR, SERPINB1, SIGLEC9
M6.14	40	Coagulation / Platelets	CD9, CD151, THBS1
M6.15	36	T Cells	CD8B, T cell receptor-like sequences
M6.18	18	Erythrocyte Development	GYPA, HEMGN, RHD, SPTB, TSPAN7, CTSE
M6.19	33	T Cells	IL21R, IL23A

Appendix C – Functional interpretation of 44 over-expressed DC modules

Module ID	Annotation	Key Transcripts	PAM3	Zymosan	Poly IC	LPS	Flagellin	R837	R848	CpG	MDP	RIG-I-L	TNFa	IFNa	IL-1b	IL-10	aCD40	Flu	Gram-	Gram+
M3.4	T Cell Activation	CCL19, IL23A				■			■										■	■
M4.14	IFN Response	IFIT1, IFIT3			■	■	■		■			■		■				■	■	
M4.15	IFN Response	INDO, IL15			■	■	■	■	■			■		■				■	■	■
M4.20	Inflammatory Response	SERPINA1		■		■			■										■	■
M4.24	T and NK Cell Activation	IL12p35, IL15, CD44				■													■	
M4.30	Inflammatory Response	SOD2, TNFSF8	■	■		■			■										■	■
M5.11	Inflammatory Response	IL1A, IL1B, IL6, TNF	■	■		■		■	■		■		■		■		■	■	■	■
M5.14	CD40 Signaling	CD40, IL15, TRAF1			■	■			■									■	■	
M5.16	Regulation of Inflammation	IL10, IL10RA, IL4R	■	■		■			■										■	■
M5.23	Regulation of inflammation	IL27, STAT4	■	■		■		■	■									■	■	■
M5.26	TLR4 Signaling	IRAK2, WTAP, F3	■	■		■			■				■		■				■	■
M5.29	IFN Response	DDX60, IFI44			■	■	■		■			■		■				■	■	
M5.33	IFN Response	IFNB1, STAT2			■	■		■	■			■		■				■	■	
M5.35	Inflammatory Response	NFKB1, IL10, CD93		■		■			■										■	■
M5.43	Inflammatory Response / ROS	NCF1, CCL5		■	■	■			■		■		■	■				■	■	■
M5.49	IFN Response	IRF1	■	■		■	■	■	■		■		■		■		■	■	■	■
M5.50	Antigen Processing/Presentation	MICB, LAG3			■	■								■				■		
M5.52	Antibacterial Response	MUC1, S100A9, SERPINB2	■	■		■			■		■								■	■
M5.53	IFN Response	BAK1, DEFB1			■									■				■		
M5.54	Activation / Maturation	CD86 , MAP4K4				■													■	
M5.56	Anti-viral Response	IFIH1, BCL2A1, TRAF1	■	■		■			■								■		■	■
M5.57	Anti-viral Response	APOBEC3G			■	■			■			■		■				■	■	
M5.61	IFN Response	CXCL10, DHX58			■	■			■			■		■				■	■	
M6.12	IFN Response / MHC Class I	HLA-E, HLA-F, IFNA10			■													■		
M6.15	Leukocyte Migration	CD47, ICAM2, TICAM2			■															
M6.16	Inflammatory Response	IL8, LILRA3, LILRA5, MIP2A	■	■		■			■									■	■	■
M6.17	IFN Response	STAT1			■	■			■					■				■	■	
M6.21	IFN Response	IFNA1, IFNW1																■		
M6.25	Cell Maturation	CCR7 , HIF1A , IL29				■			■										■	■
M6.26	IFN Response	IRF7, APOBEC3F			■	■			■			■		■				■	■	
M6.27	Inflammatory Response	IRAK3 , PLAUR , OSM	■	■					■										■	■
M6.28	IL1 signaling	NFKB2, PANX1, STARD5	■	■		■			■										■	■
M6.30	Regulation of Apoptosis	CASP3 , BCL2 , HCK																		
M6.35	Type I IFN Response	IFNAs				■			■									■	■	
M6.43	Regulation of Inflammation	SOCS1, LILRA5, STAT5A	■	■		■			■										■	■
M6.51	Inflammatory Response	NFKB, IL1RN	■	■		■			■		■		■		■				■	■
M6.52	IFN Response	IFIT5, CXCL11			■	■	■	■	■			■		■				■	■	
M6.56	Antibacterial Response	TNFRSF18, HAMP	■	■					■										■	■
M6.57	TLR8 Signaling	TLR8, TGFA	■	■					■										■	■
M6.61	IFN Response	UNC93B1			■													■		
M6.62	Inflammatory Response	IL2RA, CXCL1	■	■		■		■	■				■		■		■		■	■
M6.66	Antigen Processing	TAP2			■									■				■	■	
M6.67	Granulocyte Chemottraction	CXCL6	■	■											■				■	■
M6.8	T Cell Activation	CD70, IL12B, CCL23, HLA				■													■	

Green squares represent over-expression of the module by at least 30% in response to the corresponding stimulus in dendritic cells (populations averaged).

BIBLIOGRAPHY

1. Janeway, C.A., Jr. Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harb Symp Quant Biol* **54 Pt 1**, 1-13 (1989).
2. Akira, S., Takeda, K. & Kaisho, T. Toll-like receptors: critical proteins linking innate and acquired immunity. *Nat Immunol* **2**, 675-680 (2001).
3. Janeway, C.A., Jr. & Medzhitov, R. Innate immune recognition. *Annu Rev Immunol* **20**, 197-216 (2002).
4. Akira, S., Uematsu, S. & Takeuchi, O. Pathogen recognition and innate immunity. *Cell* **124**, 783-801 (2006).
5. Kawai, T. & Akira, S. Innate immune recognition of viral infection. *Nat Immunol* **7**, 131-137 (2006).
6. Figdor, C.G., van Kooyk, Y. & Adema, G.J. C-type lectin receptors on dendritic cells and Langerhans cells. *Nat Rev Immunol* **2**, 77-84 (2002).
7. Fritz, J.H., Ferrero, R.L., Philpott, D.J. & Girardin, S.E. Nod-like proteins in immunity, inflammation and disease. *Nat Immunol* **7**, 1250-1257 (2006).
8. Kim, T.*, et al.* Aspartate-glutamate-alanine-histidine box motif (DEAH)/RNA helicase A helicases sense microbial DNA in human plasmacytoid dendritic cells. *Proc Natl Acad Sci U S A* **107**, 15181-15186 (2010).
9. Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J.M. & Hoffmann, J.A. The dorsoventral regulatory gene cassette spatzle/Toll/cactus controls the potent antifungal response in Drosophila adults. *Cell* **86**, 973-983 (1996).
10. Kawai, T. & Akira, S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nat Immunol* **11**, 373-384 (2010).
11. Medzhitov, R., Preston-Hurlburt, P. & Janeway, C.A., Jr. A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. *Nature* **388**, 394-397 (1997).
12. Poltorak, A.*, et al.* Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science* **282**, 2085-2088 (1998).
13. Takeuchi, O.*, et al.* Differential roles of TLR2 and TLR4 in recognition of gram-negative and gram-positive bacterial cell wall components. *Immunity* **11**, 443-451 (1999).
14. Schwandner, R., Dziarski, R., Wesche, H., Rothe, M. & Kirschning, C.J. Peptidoglycan- and lipoteichoic acid-induced cell activation is mediated by toll-like receptor 2. *J Biol Chem* **274**, 17406-17409 (1999).

15. Hayashi, F.*, et al.* The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature* **410**, 1099-1103 (2001).
16. Alexopoulou, L., Holt, A.C., Medzhitov, R. & Flavell, R.A. Recognition of double-stranded RNA and activation of NF-kappaB by Toll-like receptor 3. *Nature* **413**, 732-738 (2001).
17. Lund, J.M.*, et al.* Recognition of single-stranded RNA viruses by Toll-like receptor 7. *Proc Natl Acad Sci U S A* **101**, 5598-5603 (2004).
18. Heil, F.*, et al.* Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. *Science* **303**, 1526-1529 (2004).
19. Bauer, S.*, et al.* Human TLR9 confers responsiveness to bacterial DNA via species-specific CpG motif recognition. *Proc Natl Acad Sci U S A* **98**, 9237-9242 (2001).
20. Cao, W. & Liu, Y.J. Innate immune functions of plasmacytoid dendritic cells. *Curr Opin Immunol* **19**, 24-30 (2007).
21. Muzio, M.*, et al.* Differential expression and regulation of toll-like receptors (TLR) in human leukocytes: selective expression of TLR3 in dendritic cells. *J Immunol* **164**, 5998-6004 (2000).
22. Zarember, K.A. & Godowski, P.J. Tissue expression of human Toll-like receptors and differential regulation of Toll-like receptor mRNAs in leukocytes in response to microbes, their products, and cytokines. *J Immunol* **168**, 554-561 (2002).
23. Imanishi, T.*, et al.* Cutting edge: TLR2 directly triggers Th1 effector functions. *J Immunol* **178**, 6715-6719 (2007).
24. Sallusto, F., Cella, M., Danieli, C. & Lanzavecchia, A. Dendritic cells use macropinocytosis and the mannose receptor to concentrate macromolecules in the major histocompatibility complex class II compartment: downregulation by cytokines and bacterial products. *J Exp Med* **182**, 389-400 (1995).
25. Mahnke, K.*, et al.* The dendritic cell receptor for endocytosis, DEC-205, can recycle and enhance antigen presentation via major histocompatibility complex class II-positive lysosomal compartments. *J Cell Biol* **151**, 673-684 (2000).
26. Geijtenbeek, T.B.*, et al.* Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses. *Cell* **100**, 575-585 (2000).
27. Dzionek, A.*, et al.* BDCA-2, a novel plasmacytoid dendritic cell-specific type II C-type lectin, mediates antigen capture and is a potent inhibitor of interferon alpha/beta induction. *J Exp Med* **194**, 1823-1834 (2001).
28. Willment, J.A., Gordon, S. & Brown, G.D. Characterization of the human beta -glucan receptor and its alternatively spliced isoforms. *J Biol Chem* **276**, 43818-43823 (2001).

29. Bates, E.E., *et al.* APCs express DCIR, a novel C-type lectin surface receptor containing an immunoreceptor tyrosine-based inhibitory motif. *J Immunol* **163**, 1973-1983 (1999).
30. Mehta, J.L. & Li, D. Identification, regulation and function of a novel lectin-like oxidized low-density lipoprotein receptor. *J Am Coll Cardiol* **39**, 1429-1435 (2002).
31. Valladeau, J., *et al.* Langerin, a novel C-type lectin specific to Langerhans cells, is an endocytic receptor that induces the formation of Birbeck granules. *Immunity* **12**, 71-81 (2000).
32. Matsumoto, M., *et al.* A novel LPS-inducible C-type lectin is a transcriptional target of NF-IL6 in macrophages. *J Immunol* **163**, 5039-5048 (1999).
33. Valladeau, J., *et al.* Immature human dendritic cells express asialoglycoprotein receptor isoforms for efficient receptor-mediated endocytosis. *J Immunol* **167**, 5767-5774 (2001).
34. Garcia-Vallejo, J.J. & van Kooyk, Y. Endogenous ligands for C-type lectin receptors: the true regulators of immune homeostasis. *Immunol Rev* **230**, 22-37 (2009).
35. Geijtenbeek, T.B., *et al.* DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell* **100**, 587-597 (2000).
36. Brown, G.D., *et al.* Dectin-1 is a major beta-glucan receptor on macrophages. *J Exp Med* **196**, 407-412 (2002).
37. Geijtenbeek, T.B. & Gringhuis, S.I. Signalling through C-type lectin receptors: shaping immune responses. *Nat Rev Immunol* **9**, 465-479 (2009).
38. Lambert, A.A., Imbeault, M., Gilbert, C. & Tremblay, M.J. HIV-1 induces DCIR expression in CD4+ T cells. *PLoS Pathog* **6**, e1001188 (2010).
39. Unger, W.W. & van Kooyk, Y. 'Dressed for success' C-type lectin receptors for the delivery of glyco-vaccines to dendritic cells. *Curr Opin Immunol* **23**, 131-137 (2011).
40. Riboldi, E., *et al.* Human C-type lectin domain family 4, member C (CLEC4C/BDCA-2/CD303) is a receptor for asialo galactosyl-oligosaccharides. *J Biol Chem* (2011).
41. Jahn, P.S., Zanker, K.S., Schmitz, J. & Dzionek, A. BDCA-2 signaling inhibits TLR-9-agonist-induced plasmacytoid dendritic cell activation and antigen presentation. *Cell Immunol* **265**, 15-22 (2010).
42. Goodridge, H.S. & Underhill, D.M. Fungal Recognition by TLR2 and Dectin-1. *Handb Exp Pharmacol*, 87-109 (2008).
43. Meyer-Wentrup, F., *et al.* DCIR is endocytosed into human dendritic cells and inhibits TLR8-mediated cytokine production. *J Leukoc Biol* **85**, 518-525 (2009).

44. Werts, C., Girardin, S.E. & Philpott, D.J. TIR, CARD and PYRIN: three domains for an antimicrobial triad. *Cell Death Differ* **13**, 798-815 (2006).
45. Chamaillard, M.*, et al.* An essential role for NOD1 in host recognition of bacterial peptidoglycan containing diaminopimelic acid. *Nat Immunol* **4**, 702-707 (2003).
46. Girardin, S.E.*, et al.* Nod1 detects a unique muropeptide from gram-negative bacterial peptidoglycan. *Science* **300**, 1584-1587 (2003).
47. Girardin, S.E.*, et al.* Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection. *J Biol Chem* **278**, 8869-8872 (2003).
48. Kufer, T.A., Fritz, J.H. & Philpott, D.J. NACHT-LRR proteins (NLRs) in bacterial infection and immunity. *Trends Microbiol* **13**, 381-388 (2005).
49. Dufner, A., Pownall, S. & Mak, T.W. Caspase recruitment domain protein 6 is a microtubule-interacting protein that positively modulates NF-kappaB activation. *Proc Natl Acad Sci U S A* **103**, 988-993 (2006).
50. Li, L.*, et al.* TRIP6 is a RIP2-associated common signaling component of multiple NF-kappaB activation pathways. *J Cell Sci* **118**, 555-563 (2005).
51. Viala, J.*, et al.* Nod1 responds to peptidoglycan delivered by the Helicobacter pylori cag pathogenicity island. *Nat Immunol* **5**, 1166-1174 (2004).
52. Watanabe, T.*, et al.* NOD1 contributes to mouse host defense against Helicobacter pylori via induction of type I IFN and activation of the ISGF3 signaling pathway. *J Clin Invest* **120**, 1645-1662 (2010).
53. Silva, G.K.*, et al.* Cutting edge: nucleotide-binding oligomerization domain 1-dependent responses account for murine resistance against Trypanosoma cruzi infection. *J Immunol* **184**, 1148-1152 (2010).
54. Ferwerda, G.*, et al.* NOD2 and toll-like receptors are nonredundant recognition systems of Mycobacterium tuberculosis. *PLoS Pathog* **1**, 279-285 (2005).
55. Kobayashi, K.S.*, et al.* Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. *Science* **307**, 731-734 (2005).
56. Shaw, M.H.*, et al.* T cell-intrinsic role of Nod2 in promoting type 1 immunity to Toxoplasma gondii. *Nat Immunol* **10**, 1267-1274 (2009).
57. Sabbah, A.*, et al.* Activation of innate immune antiviral responses by Nod2. *Nat Immunol* **10**, 1073-1080 (2009).
58. Moore, C.B.*, et al.* NLRX1 is a regulator of mitochondrial antiviral immunity. *Nature* **451**, 573-577 (2008).
59. Tattoli, I.*, et al.* NLRX1 is a mitochondrial NOD-like receptor that amplifies NF-kappaB and JNK pathways by inducing reactive oxygen species production. *EMBO Rep* **9**, 293-300 (2008).
60. Abdul-Sater, A.A.*, et al.* Enhancement of reactive oxygen species production and chlamydial infection by the mitochondrial Nod-like family member NLRX1. *J Biol Chem* **285**, 41637-41645 (2010).

61. Schroder, K. & Tschopp, J. The inflammasomes. *Cell* **140**, 821-832 (2010).
62. Agostini, L.*, et al.* NALP3 forms an IL-1beta-processing inflammasome with increased activity in Muckle-Wells autoinflammatory disorder. *Immunity* **20**, 319-325 (2004).
63. Martinon, F., Agostini, L., Meylan, E. & Tschopp, J. Identification of bacterial muramyl dipeptide as activator of the NALP3/cryopyrin inflammasome. *Curr Biol* **14**, 1929-1934 (2004).
64. Mariathasan, S.*, et al.* Cryopyrin activates the inflammasome in response to toxins and ATP. *Nature* **440**, 228-232 (2006).
65. Gross, O.*, et al.* Syk kinase signalling couples to the Nlrp3 inflammasome for anti-fungal host defence. *Nature* **459**, 433-436 (2009).
66. Kanneganti, T.D.*, et al.* Critical role for Cryopyrin/Nalp3 in activation of caspase-1 in response to viral infection and double-stranded RNA. *J Biol Chem* **281**, 36560-36568 (2006).
67. Muruve, D.A.*, et al.* The inflammasome recognizes cytosolic microbial and host DNA and triggers an innate immune response. *Nature* **452**, 103-107 (2008).
68. Boyden, E.D. & Dietrich, W.F. Nalp1b controls mouse macrophage susceptibility to anthrax lethal toxin. *Nat Genet* **38**, 240-244 (2006).
69. Faustin, B.*, et al.* Reconstituted NALP1 inflammasome reveals two-step mechanism of caspase-1 activation. *Mol Cell* **25**, 713-724 (2007).
70. Poyet, J.L.*, et al.* Identification of Ipaf, a human caspase-1-activating protein related to Apaf-1. *J Biol Chem* **276**, 28309-28313 (2001).
71. Mariathasan, S.*, et al.* Differential activation of the inflammasome by caspase-1 adaptors ASC and Ipaf. *Nature* **430**, 213-218 (2004).
72. Suzuki, T.*, et al.* Differential regulation of caspase-1 activation, pyroptosis, and autophagy via Ipaf and ASC in Shigella-infected macrophages. *PLoS Pathog* **3**, e111 (2007).
73. Amer, A.*, et al.* Regulation of Legionella phagosome maturation and infection through flagellin and host Ipaf. *J Biol Chem* **281**, 35217-35223 (2006).
74. Franchi, L.*, et al.* Critical role for Ipaf in Pseudomonas aeruginosa-induced caspase-1 activation. *Eur J Immunol* **37**, 3030-3039 (2007).
75. Miao, E.A., Ernst, R.K., Dors, M., Mao, D.P. & Aderem, A. Pseudomonas aeruginosa activates caspase 1 through Ipaf. *Proc Natl Acad Sci U S A* **105**, 2562-2567 (2008).
76. Molofsky, A.B.*, et al.* Cytosolic recognition of flagellin by mouse macrophages restricts Legionella pneumophila infection. *J Exp Med* **203**, 1093-1104 (2006).
77. Wright, E.K.*, et al.* Naip5 affects host susceptibility to the intracellular pathogen Legionella pneumophila. *Curr Biol* **13**, 27-36 (2003).

78. Ren, T., Zamboni, D.S., Roy, C.R., Dietrich, W.F. & Vance, R.E. Flagellin-deficient Legionella mutants evade caspase-1- and Naip5-mediated macrophage immunity. *PLoS Pathog* **2**, e18 (2006).
79. Lightfield, K.L.*, et al.* Critical function for Naip5 in inflammasome activation by a conserved carboxy-terminal domain of flagellin. *Nat Immunol* **9**, 1171-1178 (2008).
80. Yoneyama, M.*, et al.* The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nat Immunol* **5**, 730-737 (2004).
81. Kang, D.C.*, et al.* mda-5: An interferon-inducible putative RNA helicase with double-stranded RNA-dependent ATPase activity and melanoma growth-suppressive properties. *Proc Natl Acad Sci U S A* **99**, 637-642 (2002).
82. Barral, P.M.*, et al.* Functions of the cytoplasmic RNA sensors RIG-I and MDA-5: key regulators of innate immunity. *Pharmacol Ther* **124**, 219-234 (2009).
83. Hornung, V.*, et al.* 5'-Triphosphate RNA is the ligand for RIG-I. *Science* **314**, 994-997 (2006).
84. Pichlmair, A.*, et al.* RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates. *Science* **314**, 997-1001 (2006).
85. Gitlin, L.*, et al.* Essential role of mda-5 in type I IFN responses to polyriboinosinic:polyribocytidylic acid and encephalomyocarditis picornavirus. *Proc Natl Acad Sci U S A* **103**, 8459-8464 (2006).
86. Kato, H.*, et al.* Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature* **441**, 101-105 (2006).
87. Jamieson, A.M., Yu, S., Annicelli, C.H. & Medzhitov, R. Influenza virus-induced glucocorticoids compromise innate host defense against a secondary bacterial infection. *Cell Host Microbe* **7**, 103-114 (2010).
88. Guarda, G.*, et al.* Type I interferon inhibits interleukin-1 production and inflammasome activation. *Immunity* **34**, 213-223 (2011).
89. Schena, M., Shalon, D., Davis, R.W. & Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467-470 (1995).
90. Pease, A.C.*, et al.* Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc Natl Acad Sci U S A* **91**, 5022-5026 (1994).
91. Freeman, W.M., VanGuilder, H.D. & Vrana, K.E. Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques* **44**, 619-626 (2008).
92. Geiss, G.K.*, et al.* Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol* **26**, 317-325 (2008).
93. Kawasaki, E.S. The end of the microarray Tower of Babel: will universal standards lead the way? *J Biomol Tech* **17**, 200-206 (2006).

94. Jenner, R.G. & Young, R.A. Insights into host responses against pathogens from transcriptional profiling. *Nat Rev Microbiol* **3**, 281-294 (2005).
95. Der, S.D., Zhou, A., Williams, B.R. & Silverman, R.H. Identification of genes differentially regulated by interferon alpha, beta, or gamma using oligonucleotide arrays. *Proc Natl Acad Sci U S A* **95**, 15623-15628 (1998).
96. Zhu, H., Cong, J.P., Mamtora, G., Gingeras, T. & Shenk, T. Cellular gene expression altered by human cytomegalovirus: global monitoring with oligonucleotide arrays. *Proc Natl Acad Sci U S A* **95**, 14470-14475 (1998).
97. Vahey, M.T.*, et al.* Impact of viral infection on the gene expression profiles of proliferating normal human peripheral blood mononuclear cells infected with HIV type 1 RF. *AIDS Res Hum Retroviruses* **18**, 179-192 (2002).
98. Feezor, R.J.*, et al.* Molecular characterization of the acute inflammatory response to infections with gram-negative versus gram-positive bacteria. *Infect Immun* **71**, 5803-5813 (2003).
99. Kobayashi, S.D.*, et al.* Bacterial pathogens modulate an apoptosis differentiation program in human neutrophils. *Proc Natl Acad Sci U S A* **100**, 10948-10953 (2003).
100. Zhang, X.*, et al.* Gene expression in mature neutrophils: early responses to inflammatory stimuli. *J Leukoc Biol* **75**, 358-372 (2004).
101. Wang, Z.M., Liu, C. & Dziarski, R. Chemokines are the main proinflammatory mediators in human monocytes activated by Staphylococcus aureus, peptidoglycan, and endotoxin. *J Biol Chem* **275**, 20260-20267 (2000).
102. Jiang, H., Van De Ven, C., Satwani, P., Baxi, L.V. & Cairo, M.S. Differential gene expression patterns by oligonucleotide microarray of basal versus lipopolysaccharide-activated monocytes from cord blood versus adult peripheral blood. *J Immunol* **172**, 5870-5879 (2004).
103. Messmer, D., Messmer, B. & Chiorazzi, N. The global transcriptional maturation program and stimuli-specific gene expression profiles of human myeloid dendritic cells. *Int Immunol* **15**, 491-503 (2003).
104. Ju, X.S. & Zenke, M. Gene expression profiling of dendritic cells by DNA microarrays. *Immunobiology* **209**, 155-161 (2004).
105. Huang, Q.*, et al.* The plasticity of dendritic cell responses to pathogens and their components. *Science* **294**, 870-875 (2001).
106. Chaussabel, D.*, et al.* Unique gene expression profiles of human macrophages and dendritic cells to phylogenetically distinct parasites. *Blood* **102**, 672-681 (2003).
107. Nau, G.J.*, et al.* Human macrophage activation programs induced by bacterial pathogens. *Proc Natl Acad Sci U S A* **99**, 1503-1508 (2002).

108. Nau, G.J., Schlesinger, A., Richmond, J.F. & Young, R.A. Cumulative Toll-like receptor activation in human macrophages treated with whole bacteria. *J Immunol* **170**, 5203-5209 (2003).
109. Belcher, C.E.*, et al.* The transcriptional responses of respiratory epithelial cells to Bordetella pertussis reveal host defensive and pathogen counter-defensive strategies. *Proc Natl Acad Sci U S A* **97**, 13847-13852 (2000).
110. Ichikawa, J.K.*, et al.* Interaction of pseudomonas aeruginosa with epithelial cells: identification of differentially regulated genes by expression microarray analysis of human cDNAs. *Proc Natl Acad Sci U S A* **97**, 9659-9664 (2000).
111. Coombes, B.K. & Mahony, J.B. cDNA array analysis of altered gene expression in human endothelial cells in response to Chlamydia pneumoniae infection. *Infect Immun* **69**, 1420-1427 (2001).
112. Ahn, S.K., Choe, T.B. & Kwon, T.J. The gene expression profile of human umbilical vein endothelial cells stimulated with lipopolysaccharide using cDNA microarray analysis. *Int J Mol Med* **12**, 231-236 (2003).
113. Shen-Orr, S.S.*, et al.* Cell type-specific gene expression differences in complex tissues. *Nat Methods* **7**, 287-289 (2010).
114. Amit, I.*, et al.* Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* **326**, 257-263 (2009).
115. Chaussabel, D., Pascual, V. & Banchereau, J. Assessing the human immune system through blood transcriptomics. *BMC Biol* **8**, 84 (2010).
116. Bennett, L.*, et al.* Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. *J Exp Med* **197**, 711-723 (2003).
117. Baechler, E.C.*, et al.* Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc Natl Acad Sci U S A* **100**, 2610-2615 (2003).
118. Pascual, V., Allantaz, F., Arce, E., Punaro, M. & Banchereau, J. Role of interleukin-1 (IL-1) in the pathogenesis of systemic onset juvenile idiopathic arthritis and clinical response to IL-1 blockade. *J Exp Med* **201**, 1479-1486 (2005).
119. Allantaz, F.*, et al.* Blood leukocyte microarrays to diagnose systemic onset juvenile idiopathic arthritis and follow the response to IL-1 blockade. *J Exp Med* **204**, 2131-2144 (2007).
120. Allantaz, F., Chaussabel, D., Banchereau, J. & Pascual, V. Microarray-based identification of novel biomarkers in IL-1-mediated diseases. *Curr Opin Immunol* **19**, 623-632 (2007).
121. Koczan, D.*, et al.* Gene expression profiling of peripheral blood mononuclear leukocytes from psoriasis patients identifies new immune regulatory molecules. *Eur J Dermatol* **15**, 251-257 (2005).

122. Kaizer, E.C., *et al.* Gene expression in peripheral blood mononuclear cells from children with diabetes. *J Clin Endocrinol Metab* **92**, 3705-3711 (2007).
123. Takamura, T., *et al.* Gene expression profiles in peripheral blood mononuclear cells reflect the pathophysiology of type 2 diabetes. *Biochem Biophys Res Commun* **361**, 379-384 (2007).
124. Achiron, A., Feldman, A., Mandel, M. & Gurevich, M. Impaired expression of peripheral blood apoptotic-related gene transcripts in acute multiple sclerosis relapse. *Ann N Y Acad Sci* **1107**, 155-167 (2007).
125. Edwards, C.J., *et al.* Molecular profile of peripheral blood mononuclear cells from patients with rheumatoid arthritis. *Mol Med* **13**, 40-58 (2007).
126. Emamian, E.S., *et al.* Peripheral blood gene expression profiling in Sjogren's syndrome. *Genes Immun* **10**, 285-296 (2009).
127. Alcorta, D.A., *et al.* Leukocyte gene expression signatures in antineutrophil cytoplasmic autoantibody and lupus glomerulonephritis. *Kidney Int* **72**, 853-864 (2007).
128. Pascual, V., Chaussabel, D. & Banchereau, J. A genomic approach to human autoimmune diseases. *Annu Rev Immunol* **28**, 535-571 (2010).
129. Zaas, A.K., *et al.* Gene expression signatures diagnose influenza and other symptomatic respiratory viral infections in humans. *Cell Host Microbe* **6**, 207-217 (2009).
130. Ramilo, O., *et al.* Gene expression patterns in blood leukocytes discriminate patients with acute infections. *Blood* **109**, 2066-2077 (2007).
131. Pankla, R., *et al.* Genomic transcriptional profiling identifies a candidate blood biomarker signature for the diagnosis of septicemic melioidosis. *Genome Biol* **10**, R127 (2009).
132. Tang, B.M., McLean, A.S., Dawes, I.W., Huang, S.J. & Lin, R.C. Gene-expression profiling of peripheral blood mononuclear cells in sepsis. *Crit Care Med* **37**, 882-888 (2009).
133. Berry, M.P., *et al.* An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* **466**, 973-977 (2010).
134. Wu, J.Q., *et al.* Longitudinal microarray analysis of cell surface antigens on peripheral blood mononuclear cells from HIV+ individuals on highly active antiretroviral therapy. *Retrovirology* **5**, 24 (2008).
135. Tilton, J.C., *et al.* Diminished production of monocyte proinflammatory cytokines during human immunodeficiency virus viremia is mediated by type I interferons. *J Virol* **80**, 11486-11497 (2006).
136. Bosinger, S.E., *et al.* Gene expression profiling of host response in models of acute HIV infection. *J Immunol* **173**, 6858-6863 (2004).

137. Motomura, K.*, et al.* Identification of a host gene subset related to disease prognosis of HIV-1 infected individuals. *Int Immunopharmacol* **4**, 1829-1836 (2004).
138. Muffak-Granero, K.*, et al.* Study of gene expression profile in liver transplant recipients with hepatitis C virus. *Transplant Proc* **40**, 2971-2974 (2008).
139. Taylor, M.W.*, et al.* Changes in gene expression during pegylated interferon and ribavirin therapy of chronic hepatitis C virus distinguish responders from nonresponders to antiviral therapy. *J Virol* **81**, 3391-3401 (2007).
140. Nascimento, E.J.*, et al.* Gene expression profiling during early acute febrile stage of dengue infection can predict the disease outcome. *PLoS One* **4**, e7892 (2009).
141. Segal, E., Friedman, N., Kaminski, N., Regev, A. & Koller, D. From signatures to models: understanding cancer using microarrays. *Nat Genet* **37 Suppl**, S38-45 (2005).
142. Subramanian, A.*, et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).
143. Mootha, V.K.*, et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267-273 (2003).
144. Segal, E., Friedman, N., Koller, D. & Regev, A. A module map showing conditional activity of expression modules in cancer. *Nat Genet* **36**, 1090-1098 (2004).
145. Chaussabel, D.*, et al.* A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. *Immunity* **29**, 150-164 (2008).
146. Quartier, P.*, et al.* A multicentre, randomised, double-blind, placebo-controlled trial with the interleukin-1 receptor antagonist anakinra in patients with systemic-onset juvenile idiopathic arthritis (ANAJIS trial). *Ann Rheum Dis* **70**, 747-754 (2011).
147. Ogston, A. Micrococcus Poisoning. *J Anat Physiol* **17**, 24-58 (1882).
148. Klevens, R.M.*, et al.* Invasive methicillin-resistant Staphylococcus aureus infections in the United States. *Jama* **298**, 1763-1771 (2007).
149. Noskin, G.A.*, et al.* National trends in Staphylococcus aureus infection rates: impact on economic burden and mortality over a 6-year period (1998-2003). *Clin Infect Dis* **45**, 1132-1140 (2007).
150. Fleming, A. On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to Their Use in the Isolation of B. Influenzae. *Brit J Exp Pathol* **10**, 226-236 (1929).

151. Chambers, H.F. & Deleo, F.R. Waves of resistance: Staphylococcus aureus in the antibiotic era. *Nat Rev Microbiol* **7**, 629-641 (2009).
152. Barber, M. & Rozwadowska-Dowzenko, M. Infection by penicillin-resistant staphylococci. *Lancet* **2**, 641-644 (1948).
153. Barber, M. Methicillin-resistant staphylococci. *J Clin Pathol* **14**, 385-393 (1961).
154. Hiramatsu, K.*, et al.* Dissemination in Japanese hospitals of strains of Staphylococcus aureus heterogeneously resistant to vancomycin. *Lancet* **350**, 1670-1673 (1997).
155. Weigel, L.M.*, et al.* Genetic analysis of a high-level vancomycin-resistant isolate of Staphylococcus aureus. *Science* **302**, 1569-1571 (2003).
156. Sanchez Garcia, M.*, et al.* Clinical outbreak of linezolid-resistant Staphylococcus aureus in an intensive care unit. *Jama* **303**, 2260-2264 (2010).
157. Wallin, T.R., Hern, H.G. & Frazee, B.W. Community-associated methicillin-resistant Staphylococcus aureus. *Emerg Med Clin North Am* **26**, 431-455, ix (2008).
158. Crum, N.F. The emergence of severe, community-acquired methicillin-resistant Staphylococcus aureus infections. *Scand J Infect Dis* **37**, 651-656 (2005).
159. Vandenesch, F.*, et al.* Community-acquired methicillin-resistant Staphylococcus aureus carrying Panton-Valentine leukocidin genes: worldwide emergence. *Emerg Infect Dis* **9**, 978-984 (2003).
160. Carleton, H.A., Diep, B.A., Charlebois, E.D., Sensabaugh, G.F. & Perdreau-Remington, F. Community-adapted methicillin-resistant Staphylococcus aureus (MRSA): population dynamics of an expanding community reservoir of MRSA. *J Infect Dis* **190**, 1730-1738 (2004).
161. Begier, E.M.*, et al.* A high-morbidity outbreak of methicillin-resistant Staphylococcus aureus among players on a college football team, facilitated by cosmetic body shaving and turf burns. *Clin Infect Dis* **39**, 1446-1453 (2004).
162. Chambers, H.F. The changing epidemiology of Staphylococcus aureus? *Emerg Infect Dis* **7**, 178-182 (2001).
163. Kreisel, K.M.*, et al.* USA300 methicillin-resistant Staphylococcus aureus bacteremia and the risk of severe sepsis: is USA300 methicillin-resistant Staphylococcus aureus associated with more severe infections? *Diagn Microbiol Infect Dis* **70**, 285-290 (2011).
164. Miller, L.G.*, et al.* Necrotizing fasciitis caused by community-associated methicillin-resistant Staphylococcus aureus in Los Angeles. *N Engl J Med* **352**, 1445-1453 (2005).

165. Francis, J.S., *et al.* Severe community-onset pneumonia in healthy adults caused by methicillin-resistant Staphylococcus aureus carrying the Panton-Valentine leukocidin genes. *Clin Infect Dis* **40**, 100-107 (2005).
166. Rubinstein, E. Staphylococcus aureus bacteraemia with known sources. *Int J Antimicrob Agents* **32 Suppl 1**, S18-20 (2008).
167. Vander Have, K.L., *et al.* Community-associated methicillin-resistant Staphylococcus aureus in acute musculoskeletal infection in children: a game changer. *J Pediatr Orthop* **29**, 927-931 (2009).
168. Dziarski, R. & Gupta, D. The peptidoglycan recognition proteins (PGRPs). *Genome Biol* **7**, 232 (2006).
169. Wang, J.E., *et al.* Peptidoglycan and lipoteichoic acid from Staphylococcus aureus induce tumor necrosis factor alpha, interleukin 6 (IL-6), and IL-10 production in both T cells and monocytes in a human whole blood model. *Infect Immun* **68**, 3965-3970 (2000).
170. Balaban, N. & Rasooly, A. Staphylococcal enterotoxins. *Int J Food Microbiol* **61**, 1-10 (2000).
171. Blomster-Hautamaa, D.A., Kreiswirth, B.N., Kornblum, J.S., Novick, R.P. & Schlievert, P.M. The nucleotide and partial amino acid sequence of toxic shock syndrome toxin-1. *J Biol Chem* **261**, 15783-15786 (1986).
172. Lowy, F.D. Staphylococcus aureus infections. *N Engl J Med* **339**, 520-532 (1998).
173. Bhakdi, S. & Tranum-Jensen, J. Alpha-toxin of Staphylococcus aureus. *Microbiol Rev* **55**, 733-751 (1991).
174. Wright, J. Staphylococcal leucocidin (Neisser-Wechsberg type) and antileucocidin. *Lancet* **1**, 1002-1004 (1936).
175. Labandeira-Rey, M., *et al.* Staphylococcus aureus Panton-Valentine leukocidin causes necrotizing pneumonia. *Science* **315**, 1130-1133 (2007).
176. Wang, R., *et al.* Identification of novel cytolytic peptides as key virulence determinants for community-associated MRSA. *Nat Med* **13**, 1510-1514 (2007).
177. Khatib, R., *et al.* Persistent Staphylococcus aureus bacteremia: incidence and outcome trends over time. *Scand J Infect Dis* **41**, 4-9 (2009).
178. Lodise, T.P., Lomaestro, B., Graves, J. & Drusano, G.L. Larger vancomycin doses (at least four grams per day) are associated with an increased incidence of nephrotoxicity. *Antimicrob Agents Chemother* **52**, 1330-1336 (2008).
179. Cruciani, M., *et al.* Penetration of vancomycin into human lung tissue. *J Antimicrob Chemother* **38**, 865-869 (1996).
180. Wergeland, H.I., Haaheim, L.R., Natas, O.B., Wesenberg, F. & Oeding, P. Antibodies to staphylococcal peptidoglycan and its peptide epitopes, teichoic acid, and lipoteichoic acid in sera from blood donors and patients with staphylococcal infections. *J Clin Microbiol* **27**, 1286-1291 (1989).

181. Thomsen, I., *et al.* Molecular distinctions exist between community-associated methicillin-resistant Staphylococcus aureus colonization and disease-associated isolates in children. *Pediatr Infect Dis J* **30**, 418-421 (2011).
182. Campbell, S.J., *et al.* Genotypic characteristics of Staphylococcus aureus isolates from a multinational trial of complicated skin and skin structure infections. *J Clin Microbiol* **46**, 678-684 (2008).
183. Chen, C., *et al.* Removing batch effects in analysis of expression microarray data: an evaluation of six batch adjustment methods. *PLoS One* **6**, e17238 (2011).
184. Ioannidis, J.P. Microarrays and molecular research: noise discovery? *Lancet* **365**, 454-455 (2005).
185. Michiels, S., Koscielny, S. & Hill, C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* **365**, 488-492 (2005).
186. Sallusto, F., Geginat, J. & Lanzavecchia, A. Central memory and effector memory T cell subsets: function, generation, and maintenance. *Annu Rev Immunol* **22**, 745-763 (2004).
187. DeLeo, F.R., Diep, B.A. & Otto, M. Host defense and pathogenesis in Staphylococcus aureus infections. *Infect Dis Clin North Am* **23**, 17-34 (2009).
188. Sugar, C.A. & James, G.M. Finding the number of clusters in a dataset: An information-theoretic approach. *Journal of the American Statistical Association* **98**, 750-763 (2003).
189. Thomas, P.D., *et al.* PANTHER: a library of protein families and subfamilies indexed by function. *Genome research* **13**, 2129-2141 (2003).
190. Hinze, C.H., *et al.* Immature cell populations and an erythropoiesis gene-expression signature in systemic juvenile idiopathic arthritis: implications for pathogenesis. *Arthritis Res Ther* **12**, R123 (2010).
191. Herrero, J., Valencia, A. & Dopazo, J. A hierarchical unsupervised growing neural network for clustering gene expression patterns. *Bioinformatics* **17**, 126-136 (2001).
192. Foster, T.J. Immune evasion by staphylococci. *Nat Rev Microbiol* **3**, 948-958 (2005).
193. Fournier, B. & Philpott, D.J. Recognition of Staphylococcus aureus by the innate immune system. *Clin Microbiol Rev* **18**, 521-540 (2005).
194. Hadley, J.S., Wang, J.E., Foster, S.J., Thiemermann, C. & Hinds, C.J. Peptidoglycan of Staphylococcus aureus upregulates monocyte expression of CD14, Toll-like receptor 2 (TLR2), and TLR4 in human blood: possible implications for priming of lipopolysaccharide signaling. *Infect Immun* **73**, 7613-7619 (2005).

195. Bueno, C., *et al.* Flow cytometric analysis of cytokine production by normal human peripheral blood dendritic cells and monocytes: comparative analysis of different stimuli, secretion-blocking agents and incubation periods. *Cytometry* **46**, 33-40 (2001).
196. Sundstedt, A., Grundstrom, S. & Dohlsten, M. T cell- and perforin-dependent depletion of B cells in vivo by staphylococcal enterotoxin A. *Immunology* **95**, 76-82 (1998).
197. Viau, M., Longo, N.S., Lipsky, P.E. & Zouali, M. Staphylococcal protein a deletes B-1a and marginal zone B lymphocytes expressing human immunoglobulins: an immune evasion mechanism. *J Immunol* **175**, 7719-7727 (2005).
198. Holub, M., *et al.* Lymphocyte subset numbers depend on the bacterial origin of sepsis. *Clin Microbiol Infect* **9**, 202-211 (2003).
199. Lanzavecchia, A. & Sallusto, F. Understanding the generation and function of memory T cell subsets. *Curr Opin Immunol* **17**, 326-332 (2005).
200. Lanzavecchia, A. & Sallusto, F. Dynamics of T lymphocyte responses: intermediates, effectors, and memory cells. *Science* **290**, 92-97 (2000).
201. Mobley, J.L., Leininger, M., Madore, S., Baginski, T.J. & Renkiewicz, R. Genetic evidence of a functional monocyte dichotomy. *Inflammation* **30**, 189-197 (2007).
202. Scumpia, P.O., *et al.* Cutting edge: bacterial infection induces hematopoietic stem and progenitor cell expansion in the absence of TLR signaling. *J Immunol* **184**, 2247-2251 (2010).
203. Nagai, T., *et al.* Timing of IFN-beta exposure during human dendritic cell maturation and naive Th cell stimulation has contrasting effects on Th1 subset generation: a role for IFN-beta-mediated regulation of IL-12 family cytokines and IL-18 in naive Th cell differentiation. *J Immunol* **171**, 5233-5243 (2003).
204. Tanno, T. & Miller, J.L. Iron Loading and Overloading due to Ineffective Erythropoiesis. *Adv Hematol* **2010**, 358283 (2010).
205. Wojcik, A.J., Skaflen, M.D., Srinivasan, S. & Hedrick, C.C. A critical role for ABCG1 in macrophage inflammation and lung homeostasis. *J Immunol* **180**, 4273-4282 (2008).
206. Gonzalez, B.E., *et al.* Venous thrombosis associated with staphylococcal osteomyelitis in children. *Pediatrics* **117**, 1673-1679 (2006).
207. Bouchoucha, S., *et al.* Deep venous thrombosis associated with acute hematogenous osteomyelitis in children. *Orthop Traumatol Surg Res* **96**, 890-893 (2010).
208. Hollmig, S.T., Copley, L.A., Browne, R.H., Grande, L.M. & Wilson, P.L. Deep venous thrombosis associated with osteomyelitis in children. *J Bone Joint Surg Am* **89**, 1517-1523 (2007).

209. Balamayooran, G., Batra, S., Fessler, M.B., Happel, K.I. & Jeyaseelan, S. Mechanisms of neutrophil accumulation in the lungs against bacteria. *Am J Respir Cell Mol Biol* **43**, 5-16 (2010).
210. Cheong, C.*, et al.* Microbial stimulation fully differentiates monocytes to DC-SIGN/CD209(+) dendritic cells for immune T cell areas. *Cell* **143**, 416-429 (2010).
211. Iijima, N., Mattei, L.M. & Iwasaki, A. Recruited inflammatory monocytes stimulate antiviral Th1 immunity in infected tissue. *Proc Natl Acad Sci U S A* **108**, 284-289 (2011).
212. Blanco, P., Palucka, A.K., Gill, M., Pascual, V. & Banchereau, J. Induction of dendritic cell differentiation by IFN-alpha in systemic lupus erythematosus. *Science* **294**, 1540-1543 (2001).
213. Hanses, F.*, et al.* Intracellular survival of Staphylococcus aureus in adipocyte-like differentiated 3T3-L1 cells is glucose dependent and alters cytokine, chemokine, and adipokine secretion. *Endocrinology* **152**, 4148-4157 (2011).